Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρονικής & Υπολογιστών

Χρήση των Generative Adversarial Networks για επιλογή πόζας και στιλ σε εφαρμογές σχεδιασμού μόδας

Διπλωματική Εργασία

του

Αθανάσιου Χαρισούδη

Επιβλέποντες: κ. Περικλής Μήτκας
Καθηγητής Α.Π.Θ., ISSEL
κ. Αντώνης Χρυσόπουλος
Μεταδιδακτορικός Ερευνητής, ISSEL

22 Ιουλίου 2021

Περίληψη

Η Παραγωγική Μοντελοποίηση αποτελεί τον κλάδο της Μηχανικής Μάθησης ο οποίος επικεντρώνεται στην παραγωγή νέων ρεαλιστικών δεδομένων και ο οποίος αποτελούσε παραδοσιακά το ανώφλι των δυνατοτήτων των μοντέλων Μηχανικής και Βαθιάς Μάθησης. Το σκηνικό έχει αλλάξει τα τελευταία χρόνια, κυρίως από το 2014, όταν ο I. Goodfellow παρουσίασε την ιδέα του για ένα παραγωγικό μοντέλο αποτελούμενο από δύο αντιμαχόμενα νευρωνικά δίκτυα, το οποίο ονομάστηκε Generative Adversarial Network. Ακολούθως, προέκυψε πληθώρα μοντέλων βασιζόμενα σε αυτό, με εντυπωσιακά αποτελέσματα που, ειδικά στο πλαίσια της παραγωγής εικόνων, εκπλήσσουνε ακόμα και ένα έμπειρο ανθρώπινο σύστημα όρασης.

Ταυτόχρονα, τις τελευταίες δεκαετίες, γίνεται όλο και περισσότερη έρευνα γύρω από την ανάπτυξη τεχνικών κατανόησης της έννοιας της μόδας και της τάσης αυτής, είτε για συνδρομή κατά τη φάση σχεδίασης ρούχων ή για καλύτερες και πιο στοχευμένες αγορές ή για άλλους σκοπούς. Στην προσπάθειά μας να εφαρμόσουμε σύγχρονες τεχνικές μηχανικής μάθησης για αυτοματοποιημένη παραγωγή και επεξεργασία εικόνων μόδας, στην παρούσα εργασία, χρησιμοποιούμε Generative Adversarial Networks. Συγκεκριμένα, σχεδιάζουμε και υλοποιούμε ένα πολυ-εργαλείο αυτοματοποιημένης επεξεργασίας εικόνων μόδας το οποίο οπλίζουμε με τέσσερις (4) βασικές λειτουργίες: αλλαγή πόζας, εξαγωγή ρούχου, ταίριασμα στιλ και παραγωγή ρεαλιστικών εικόνων κατ΄ απαίτηση.

Στην προσπάθειά μας αυτή, εκπαιδεύουμε ισάριθμα μοντέλα τύπου Generative Adversarial Network σε σύνολα δεδομένων αποτελούμενα από εικόνες μόδας (ενν. ρούχων και μοντέλων που τα διαφημίζουν), παραθέτοντας στο τέλος σχετικά αποτελέσματα. Αποτελεί βαθιά πεποίθησή μας ότι εξελίξεις παρόμοιων μοντέλων θα κατέχουν κεντρικό ρόλο στη σχεδίαση ρούχων και κυρίως στη διάθεσή τους μέσω συστημάτων ηλεκτρονικού εμπορίου στο εγγύς μέλλον, κάτι που μας έκανε να προσηλωθούμε με ζήλο στην υλοποίηση ενός αποτελεσματικού νοήμονος εργαλείου για επεξεργασία εικόνων μόδας στην παρούσα εργασία.

Λέξεις κλειδιά— Παραγωγική Μοντελοποίηση, Νοήμονα συστήματα μόδας, Generative Adversarial Networks, Παραγωγή εικόνων από θόρυβο, Μετατροπή εικόνας-σε-εικόνα, DeepFashion, StyleGAN, CycleGAN, Μηχανική Όραση, Τεχνητά Νευρωνικά Δίκτυα

Abstract

Generative Modelling, a branch of Machine Learning that focuses on generating realistic-looking samples, has traditionally constituted the upper bound of what Machine and Deep Learning models can achieve. This regime has completely changed the past years, especially after 2014, when I. Goodfellow presented his idea for a generative model comprising two competing neural networks: the Generative Adversarial Network of GAN for short. Subsequently, a plethora of models based on GAN have been proposed with impressive results, some of which, principally in the context of image generation, surprise even an experienced human vision system.

Concurrently, more and more research is devoted during the last decades around the development of techniques for demystifying the notion of fashion and fashion trends. Among its purposes, is creating artificial intelligence systems that provide help in the process of designing new garments as well as in the process of conducting better and more well-targeted purchases. In an endeavour to apply modern machine learning techniques to automate generation and editing of fashion images, in this project we employ Generative Adversarial Networks. In particular, we design and utilize a multi-tool for automatic editing of fashion images, equipped with four (4) fundamental operations: pose change, cloth extraction, style matching and on-demand realistic fashion images generation.

In order to achieve our goals, we train four models based on the Generative Adversarial Network in fashion image (i.e. images of garments as well as human models advertising them) datasets, giving the corresponding outcomes at the end. It is our firm belief that further developments of such models will play a central role in fashion design and especially in clothes distribution through e-commerce systems in the near future, which has made us focus zealously on implementing an effective intelligent tool for fashion image editing in this work.

Keywords— Generative Modelling, Intelligent fashion systems, Generative Adversarial Networks, Noise-to-image generation, Image-to-image translation, DeepFashion, StyleGAN, CycleGAN, Computer Vision, Artificial Neural Networks

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τον κ. Μήτκα Περικλή και τον κ. Αντώνη Χρυσόπουλο για την εμπιστοσύνη που μου έδειξαν τόσο κατά την ανάθεση της παρούσας διπλωματικής εργασίας όσο και κατά την εκπόνησή της. Επίσης, τους ευχαριστώ θερμά για την αμέριστη υποστήριξη και καθοδήγηση που μου παρείχαν έως και την τελευταία ημέρα εκπόνησης αυτής.

Επίσης, θέλω να ευχαριστήσω τον κ. Αλέξανδρο Κυπριανίδη και τον κ. Σωτήρη Τσαρούχη για την πολύτιμη βοήθεια που μου παρείχαν, και τη γενικότερη συμβουλευτική τους παρουσία.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, την αδελφή και τον αδελφό μου, που είναι πάντα δίπλα μου και με στηρίζουν σε όλη τη διάρκεια της ζωής μου, καθώς και τους φίλους μου, τους συμφοιτητές μου και ένα δικό μου πρόσωπο, που γέμισαν τα φοιτητικά μου χρόνια με αξέχαστες εμπειρίες.

Χρήση των Generative Adversarial Networks για επιλογή πόζας και στιλ σε εφαρμογές σχεδιασμού μόδας

Αθανάσιος Χαρισούδης
achariso@ece.auth.gr

22 Ιουλίου 2021

Περιεχόμενα

Katˆlogoc Pinˆkwn

Katˆlogoc Sqhmˆtwn

Kefˆlaio 1

Sqetikˆ me th Diplwmatik ErgasÐa

AntikeÐmeno kai Skopìc

Sthn paroÔsa diplwmatik ergasÐa asqoloÔmaste me Paragwgik MontelopoÐhsh eikìnwn, dhlad prospajoÔme na ekpaideÔsoume montèla Mhqanik c/Bajiˆc Mˆjhshc ¸ste na parˆgoun realistikèc eikìnec. Gia na to petÔqoume autì, sqediˆzoume ta montèla me tètoio trìpo ¸ste na mporoÔn na afomoi¸soun ta basikˆ qarakthristikˆ kai th dom twn eikìnwn enìc sunìlou dedomènwn ekpaÐdeushc. Sta plaÐsia autˆ, qrhsimopoioÔme Bajiˆ Paragwgikˆ Montèla tÔpou Generative Adversarial Network (GAN), ta opoÐa apoteloÔntai apì dÔo neurwnikˆ dÐktua kai ekpaideÔontai antiparajetikˆ: ìso to èna dÐktuo gÐnetai kalÔtero tìso to ˆllo metabˆllei tic paramètrouc tou gia na anakt sei thn kuriarqÐa.

Çtsi, sqediˆsame kai ulopoi same montèla GAN, ta opoÐa ekpaideÔsame se sÔnola dedomènwn apoteloÔmena apì eikìnec mìdac. Se mÐa genik eikìna, autì pou prospaj same na ulopoi soume sta plaÐsia thc paroÔsac diplwmatik c ergasÐac eÐnai èna poluergaleÐo mhqanik c mˆjhshc me dunatìthtec paragwg c realistik¸n eikìnwn mìdac allˆ kai efarmog c realistik¸n metasqhmatism¸n stic eikìnec autèc . To ergaleÐo autì to oplÐsame me tic akìloujec dunatìthtec:

1. Allag Pìzac : dojeÐshc mÐac eikìnac eisìdou kai miac pìzac thc eikìnac exìdou, to antÐstoiqo montèlo tou ergaleÐou mac prospajeÐ na parˆxei mÐa realistik eikìna pou na moiˆzei ìso to dunatìn perissìtero sthn pragmatik eikìna exìdou (enn. me th nèa pìza)

2. Exagwg RoÔqou: dojeÐshc miac eikìnac enìc anjr¸pou (se aujaÐreto parask nio), to antÐstoiqo montèlo tou ergaleÐou mac prospajeÐ na parˆxei mÐa eikìna sthn opoÐa na apeikonÐzetai mìno to roÔqo tou anjr¸pou se oudètero parask nio

3. TaÐriasma Stil : dojeÐshc mÐac eikìnac upod matoc (antÐstrofa mÐac tsˆntac qerioÔ), to antÐstoiqo montèlo tou ergaleÐou mac prospajeÐ na parˆxei mÐa eikìna tsˆnta (antÐstrofa upod matoc) h opoÐa na tairiˆzei stilistikˆ (p.q. sto qr¸ma sthn epishmìthta) sthn arqik eikìna

4. Paragwg realistik¸n eikìnwn mìdac : tèloc, oplÐsame to polu-ergaleÐo mac me èna apì ta plèon exeligmèna montèla GAN, to StyleGAN, me skopì autì na èqei th dunatìthta paragwg c realistik¸n eikìnwn mìdac ìpou apeikonÐzontai ˆnjrwpoi-montèla se diˆforec pìzec kai for¸ntac diˆfora sÔnola roÔqwn, en¸ se mellontik epèktash ja eÐnai dunat h elegqìmenh meÐxh metaxÔ twn paragìmenwn eikìnwn gia allag stil/roÔqwn/pìzac k.o.k.

Diˆrjrwsh

H ergasÐa diarjr¸jhke se pènte kefˆlaia (kefˆlaia 2-6) plhn tou parìntoc kai tou teleutaÐou kefalaÐou me mellontikèc proektˆseic. Se kˆje èna apì ta kefˆlaia autˆ perilambˆnontai ta ex c:

- Kefˆlaio 2 : sto kefˆlaio autì perilambˆnontai genikèc bibliografikèc anaforèc sqetikèc analÔseic pou eisˆgoun ton anagn¸sth sthn Paragwgik Montelopoðhsh eikìnwn kai sta Generative Adversarial Networks (GANs) pou apoteloÔn kai ton pur na thc paroÔsac ergasÐac.

- Kefˆlaio 3 : akoloujeÐ èna kefˆlaio pou estiˆzei sta GANs, se ìti aforˆ th dìm kai ton trìpo ekpaÐdeus c touc en¸ perilambˆnei kai tic katexoq n metrikèc gia axiolìghshc thc apìdos c touc.

- Kefˆlaio 4 : katìpin pernˆme stic efarmogèc twn GANs kaj¸c kai se arqitektonikèc tètoiwn montèlwn pou èqoun protajeÐ sth bibliografÐa kai èqoun qrhsimopoihjeÐ sth prˆxh sta plaÐsia Paragwgik c MontelopoÐhshc eikìnwn. EkeÐ perigrˆfoume kai ta montèla sta opoÐa basist kame gia thn ulopoÐhsh tou polu-ergaleÐou mac dÐnontac antÐstoiqa apotelèsmata apì th sqetik bibliografÐa.

- Kefˆlaio 5 : autì kai to epìmeno kefˆlaio apoteloÔn to epÐkentro thc ulopoÐhs c mac me to kefˆlaio 5 na perilambˆnei plhroforÐec gia ta sÔnola dedomènwn pou qrhsimopoi jhkan, th mejodologÐa pou akolouj jhke gia thn pro-epexergasÐa aut¸n, ta montèla pou sqediˆsthkan kai ulopoi jhkan kaj¸c kai paramètrouc ekpaÐdeushc twn montèlwn aut¸n.

- Kefˆlaio 6 : sto kefˆlaio 6 proqwroÔme sthn parˆjesh metrik¸n katˆ thn ek-paÐdeush twn montèlwn mac kaj¸c kai parˆjesh metrik¸n kai apotelesmˆtwn afìtou aut èqei oloklhrwjeÐ.

- Kefˆlaio 7 : sto sÔntomo autì kefˆlaio dÐnoume seirˆ mellontik¸n proektˆsewn pou tìso o ekponht c ìso kai o endiaferìmenoc anagn¸sthc mporoÔn na qrhsimo-poi soun gia th sunèqish thc paroÔsac diplwmatik c ergasÐac.

Prin proqwr soume, anafèroume sto shmeÐo autì pwc h ìlh anˆptuxh kai ekpaÐdeush twn montèlwn ègine sthn programmatistik gl¸ssa Python (elˆq. ekd. 3.7), en¸ qrhsi-mopoi jhke to sÔnolo bibliojhk¸n PyTorch (elaq. ekd. 1.7.2). O k¸dikac thc paroÔsac diplwmatik c ergasÐac, o opoÐoc xepernˆei tic 15.5 K LoC, dÐnetai anoiqtˆ allˆ qwrÐc eujÔnh sto akìloujo apojet rio k¸dika tou GitHub:

github.com/achariso/gans-thesis

Kefˆlaio 2

Eisagwg sta Generative Adversarial Networks (GANs) kai sthn Paragwgik MontelopoÐhsh

SÔmfwna me ton orismì thc Mhqanik c Eufuòac pou dìjhke apì ton Alan M. Turing sto ˆrjro tou Computing Machinery and Intelligence[3], aut mporeÐ na stoiqeiojethjeÐ gia mia mhqan ìtan oi apokrÐseic thc eÐnai dÔskolo na diakrijoÔn apì autèc enìc anjr¸pou (kai ètsi h mhqan ja pernoÔse apì to test gnwstì wc imitation game). Parìlo pou se kˆpoiouc tomeÐc thc Mhqanik c Mˆjhshc ìpwc h AllhlepÐdrash Fusik c Gl¸ssac (Conversational AI) upˆrqei akìma shmantik apìstash mèqri oi mhqanèc na apokrÐnontai me realistikì kai anjr¸pino trìpo, se ˆllouc tomeÐc ìpwc h Anagn¸rish ProtÔpwn kai kat' epèktash h Taxinìmhsh Prìbleyh, oi upologistikèc mhqanèc epideiknÔoun akìmh kai uperoq se sÔgkrish me touc anjr¸pouc.

Sthn paroÔsa diplwmatik ergasÐa epikentrwnìmaste sth melèth teqnik¸n kai algorÐjmwn Paragwgik c MontelopoÐhshc (Generative Modeling) sta plaÐsia thc Mhqanik c —rashc (Computer Vision), dhlad Paragwgik MontelopoÐhsh eikìnwn. H Paragwgik MontelopoÐhsh, estiˆzontac sthn paragwg teqnht¸n polumesik¸n dedomènwn pou parousiˆzoun poikilomorfÐa kai realismì, paradosiakˆ apoteloÔse thn {AqÐlleio ptèrna} thc Teqnht c NohmosÔnhc. Prìsfatec exelÐxeic, wstìso, ston tomèa thc Bajiˆc Mˆjhshc (Deep Learning), se sunduasmì me th suneq beltÐwsh tou ulikoÔ (hardware) twn upologist¸n, èqoun odhg sei thn ereunhtik koinìthta sthn anaz thsh

nèwn mejìdwn Paragwgik c MontelopoÐhshc, idiaÐtera se ìti aforˆ th dhmiourgÐa realistik¸n disdiˆstatwn kai trisdiˆstatwn apeikonÐsewn. Arqikˆ, h Paragwgik MontelopoÐhsh gia Mhqanik —rash basÐzontan stouc Autìmatouc Kwdikopoihtèc (Autoencoders), h efarmog twn opoÐwn wc endiˆmesec str¸seic (layers) exagwg c qarakthristik¸n se BajiˆNeurwnikˆ DÐktua (ìpwc sto [18]), od ghse sthn paragwg twn pr¸twn axiìlogwn kai arketˆ realistik¸n eikìnwn stic arqèc tou ai¸na.

Sq ma 1: PoikilÐa (diversity) apì deÐgmata qeirìgrafwn yhfÐwn pou èqoun paraqjeÐ apì DÐktuo Stacked Denoising Autoencoder (SDAE)me treic (3) krufèc str¸seic.
Phg : Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion , Pascal Vincent et al. , 2010 [18]

›swc h pio axioshmeÐwth prosj kh stic teqnikèc Paragwgik c MontelopoÐhshc eikìnwn rje to 2014 ìtan o Ian Goodfellow, katˆ th diˆrkeia thc didaktorik c tou diatrib c sto Panepist mio tou Montreal, efhÔre ta Paragwgikˆ AntiparajetikˆDÐktua Generative Adversarial Networks (GANs). H teqnik aut basizìmenh sthn tautìqronh ekpaÐdeush dÔo (2) xeqwrist¸n Baji¸n Neurwnik¸n DiktÔwn (DNNs), od ghse sthn paragwg eikìnwn o bajmìc poiìthtac kai realismoÔ twn opoÐwn dièkrine amèswc thn teqnik aut apì ìlec tic prohgoÔmenec. Gia parˆdeigma, h qr sh twn GANs epètreye thn ektèlesh ergasi¸n pou èwc tìte jewroÔntan adÔnatec, ìpwc thn paragwg teqnht¸n (fake) eikìnwn me poiìthta efˆmillh eikìnwn tou pragmatikoÔ kìsmou, th metatrop enìc skÐtsou se mia eikìna shmantikˆ ìmoia me fusikèc eikìnec, th metatrop enìc bÐnteo ìpou eikonÐzetai èna ˆlogo se èna ˆllo bÐnteo, ìpou to ˆlogo èqei antikatastajeÐ me zèbra • kai ìla autˆ qwrÐc thn anˆgkh Ôparxhc terˆstiwn set apì epimel¸c epishmasmèna (annotated) dedomèna ekpaÐdeushc. Sto sq ma 2 pou paratÐjetai akoloÔjwc, faÐnetai afenìc h ikanìthta twn GANs na parˆgoun realistikèc eikìnec kai afetèrou o ragdaÐoc rujmìc me ton opoÐo proqwrˆei kai anaptÔssetai h ereÔna

gÔrw apì aut thn epanastatik teqnik Paragwgik c MontelopoÐhshc. Aut prìodoc afenìc epibebai¸nei thn estÐash thc ereunhtik c koinìthtac ta teleutaÐa qrìnia sthn uiojèthsh kai anˆptuxh GANs en¸ afetèrou apoteleÐ peist rio thc uperoq c touc metaxÔ twn teqnik¸n Paragwgik c MontelopoÐhshc.

Sq ma 2: Prìodoc sthn paragwg teqnht¸n eikìnwn me prìswpa anjr¸pwn qrhsimopoi¸ntac Generative Adversarial Networks (GANs): apì asprìmaurec eikìnec to 2014 se èggrwmec fwtore-alistkèc eikìnec to 2018.
Phg : Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments , Salehi et al., 2020 [114]

Prin proqwr soume, wstìso, sthn parousÐash thc basik c dom c kai qarakthristik¸n twn GANs, jewroÔme skìpimo na orÐsoume saf¸c thn Paragwgik MontelopoÐhsh gia Mhqanik —rash qrhsimopoi¸ntac teqnikèc Bajiˆc Mˆjhshc, na sugkrÐnoume ta Diakritikˆ me ta Paragwgikˆ Montèla, kaj¸c kai na anaferjoÔme se kombikèc teqnikèc Paragwgik c MontelopoÐhshc, ìpwc ta AutopalindronoÔmena Paragwgikˆ Montèla kai oi MetablhtoÐ Autìmatoi Kwdikopoihtèc (Variational Autoencoders). Oi Variational Autoencoders (VAEs), sugkekrimèna, ousiastikˆ apoteloÔn ˆmeso prìgono twn GANs kai moirˆzontai pollˆ koinˆ ston trìpo kataskeu c kai ekpaÐdeushc me autˆ, me thn pio shmantik diaforopoÐhsh touc na entopÐzetai sto sqhmatismì thc sunˆrthsh kìstouc.

2.1 Paragwgik MontelopoÐhsh

Se epèktash thc diadedomènhc r shc tou AmerikanoÔ fusikoÔ Richard Feynman, "What I cannot create, I do not understand", ja mporoÔsame na kinhtrodot soume th dhmiourgÐa

teqnik¸n Paragwgik c MontelopoÐhshc me th frˆsh "What I understand, I can create" [100]. Mia ermhneÐa aut c sta plaÐsia thc Mhqanik c —rashc ja mporoÔse na eÐnai ìti efìson h lanjˆnousa {dom } pijanotik katanom enìc sunìlou pragmatik¸n eikìnwn mporeÐ epark¸c na {aiqmalwtisteÐ} apì kˆpoio montèlo Paragwgik c MontelopoÐhshc me Mhqanik /Bajiˆ Mˆjhsh, tìte to montèlo autì ja prèpei na eÐnai ikanì na parˆgei nèec eikìnec deÐgmata, mh-diakrÐsima apì tic eikìnec deÐgmata pou proèrqontai apì to sÔnolo dedomènwn ekpaÐdeushc.

Sthn paroÔsa ergasÐa, epikentrwnìmaste se montèla Statistik c Paragwgik c MontelopoÐhshc (Statistical Generative Modeling) me DÐktua Bajiˆc Mˆjhshc, ta opoÐa kai analÔontai akoloÔjwc. Upˆrqoun arketoÐ tÔpoi tètoiwn montèlwn pou èqoun anaptuqjeÐ sth bibliografÐa allˆ kai sth prˆxh. Çnac apì autoÔc eÐnai kai ta GANs, wstìso prin ftˆsoume sthn anˆlush thc dom c kai leitourgÐac touc, ja gÐnei anaforˆ kai se ˆllouc tÔpouc teqnik¸n kai montèlwn Statistik c Paragwgik c MontelopoÐhshc me DÐktua Bajiˆc Mˆjhshc.

Statistik Paragwgik MontelopoÐhsh

Ta montèla Statistik c Paragwgik c MontelopoÐhshc Statistikˆ Paragwgikˆ Montèla (Statistical Generative Models), eÐnai ekpaideÔsima montèla ta opoÐa basÐzontai sthn Ôparxh megˆlwn sunìlwn dedomènwn prokeimènou na {aiqmalwtÐsoun} rhtˆ èmmesa mia pijanotik katanom ($P(X)$ eˆn prìkeitai gia paragwg qwrÐc sunj kh, $P(X|Y)$ gia upo-sunj kh paragwg /katanom - kˆti pou analÔetai ekten¸c parakˆtw). Ta ekpaideumèna statistikˆ autˆ montèla kaloÔntai en suneqeÐa na parˆgoun deÐgmata pou akoloujoÔn thn Ðdia pijanotik katanom , ìpou wc X nooÔntai ta parathr sima dedomèna (p.q. mia eikìna tou sunìlou dedomènwn ekpaÐdeushc), en¸ wc Y noeÐtai h klˆsh tˆxh sthn opoÐa an kei to kˆje èna apì ta dedomèna ekpaÐdeushc kai h opoÐa qrhsimopoieÐtai apì to ekpaideumèno montèlo gia thn paragwg deigmˆtwn thc antÐstoiqhc klˆshc tˆxhc (eˆn prìkeitai gia mh-epiblepìmenh ekpaÐdeush to Y sun jwc paraleÐpetai opìte kai milˆme gia paragwg /katanom qwrÐc sunj kh).

Sth genik perÐptwsh, mac dÐnetai èna sÔnolo dedomènwn x_i pou akoloujeÐ kˆpoia katanom $p_{data}(x)$ kai stìqoc thc Statistik c Paragwgik c MontelopoÐhshc eÐnai h eÔresh enìc montèlou (dhl. arqikˆ miac oikogèneiac montèlwn kai sth sunèqeia beltistopoÐhsh gia eÔresh twn bèltistwn paramètrwn) tou opoÐou h prosèggish thc $p_{data}(x)$, $p_{model;\tilde{}}(x)$ me paramètrouc tic $\tilde{}$ ja èqei th mikrìterh dunat {apìstash} apì aut . Eˆn

san mètro thc apìstashc katanom¸n qrhsimopoihjeÐ h KL Divergence[4], tìte prokÔptei o diadedomènoc stìqoc ekpaÐdeushc Statistik¸n Paragwgik¸n Montèlwn: eÔresh tou montèlou pou megistopoieÐ th logarijmik pijanof^neia (log-likelihood) pou anatÐjetai sta dedomèna tou sunìlou ekpaÐdeushc .

Epomènwc, mporoÔme na jewr soume èna Statistikì Paragwgikì Montèlo wc mia ekpaideÔsimh sun^rthsh katanom c pijanìthtac gia thn ekpaÐdeush thc opoÐac apaitoÔntai afenìc dedomèna(data) kai afetèrou prìterh gn¸sh (prior knowledge) h opoÐa upeisèrqetai sto montèlo eÐte wc h morf thc sun^rthshc katanom c (p.q. Gaussian), eÐte wc h sun^rthsh kìstouc gia thn eÔresh thc bèltisthc lÔshc (p.q. maximum likelihood - cross entropy), eÐte wc o algìrijmoc beltistopoÐhshc pou ja qrhsimopoihjeÐ (p.q. stochastic gradient descent) wc k^ti ^llo. H an^gkh Ôparxhc prìterhc gn¸shc èrqetai wc ^mesh sunèpeia tou gnwstoÔ jewr matoc No Free Lunch Theoremsta plaÐsia thc beltistopoÐhshc kai thc baji^c m^jhshc [21][75].

SÔgkrish Diakritik¸n kai Paragwgik¸n Montèlwn

Sta plaÐsia thc Mhqanik c kai Baji^c M^jhshc, wc summetrik^ twn Paragwgik¸n Montèlwn nooÔntai ta Diakritik^ Montèla (Discriminative Models) , ta opoÐa eÐnai aut^ pou paradosiak^ qrhsimopoioÔntai se efarmogèc ìpwc h Taxinìmhsh (Classi cation) h OmadopoÐhsh(Clustering). Ta Statistik^ Diakritik^ Montèla (Statistical Discriminative Models) ekpaideÔontai se sÔnola dedomènwn me skopì na mporoÔn na diaqwrÐzoun ta ek^stote dedomèna eisìdou se di^forec t^xeic (classes) om^dec (clusters). Epomènwc, lamb^nontac dedomèna eisìdou, X, all^ kai exìdou (efîson up^rqoun), Y, kaloÔntai na {m^joun} th sun^rthsh katanom c pijanìthtac $P(Y|X)$. Autì dikaiologeÐ th qrhsimopoÐhsh tou ìrou {summetrik^} wc proc ta Paragwgik^ Montèla, ta opoÐa {majaÐnoun} thn $P(X|Y)$, k^ti pou faÐnetai kai ston pÐnaka pou akoloujeÐ.

PÐnakac 1: Omoiìthtec kai diaforèc anˆmesa sta Paragwgikˆ kai sta Diakritikˆ Montèla Mhqanik c Mˆjhshc sto plaÐsio thc Anagn¸rishc ProtÔpwn.
Phg : Supervised Learning Cheatsheet - Stanford CS 229 (Machine Learning) [124]

	Diakritikˆ Montèla	Paragwgikˆ Montèla
Stìqoc EkpaÐdeushc	'Amesh (rht) ektÐmhsh thc $P(Y\|X)$	'Amesh èmmesh ektÐmhsh thc $P(X\|Y)$
Apotèlesma EkpaÐdeushc	To —rio Apìfashc (decision boundary)	H (upì-sunj kh) sunˆrthsh puknìthtac pijanìthtac twn dedomènwn ekpaÐdeushc
OptikopoÐhsh		
ParadeÐgmata Arqitektonik¸n	Support Vector Machines [9], Conditional Random Fields [12]	Gaussian Discriminant Analysis (GDA) [2], Naive Bayes Classi er

Sta plaÐsia thc Mhqanik c —rashc me Bajiˆ Mˆjhsh, èqoun anaptuqjeÐ arketˆ kainotìmec kai apodotikèc arqitektonikèc Diakritik¸n Montèlwn. Sugkekrimèna, èqoun qrhsimopoihjeÐ ekten¸c montèla me Bajiˆ SuneliktikˆNeurwnikˆ DÐktua (Deep Convolutional Neural Networks - Deep CNNs) se efarmogèc ìpwc:

- ˆ Taxinìmhsh Eikìnac (Image Classi cation)
- ˆ Anagn¸rish Antikeimènwn se Eikìna (Object Detection)
- ˆ Apìdosh Etikètac se kˆje eikonostoiqeÐo miac Eikìnac (Instance Segmentation)
- ˆ Anˆlush - Taxinìmhsh BÐnteo (Video Analysis - Classi cation)

Pollèc apì tic kainotomÐec kai arqitektonikèc diktÔwn pou epino jhkan sta Diakritikˆ Montèla katˆ thn efarmog touc stic parapˆnw kathgorÐec me kÔria thn taxinìmhsh eikìnwn, uiojet jhkan kai kajier¸jhkan kai sta Bajiˆ Paragwgikˆ Montèla, ìpwc analÔetai se epìmenh upoenìthta. Parakˆtw, anafèroume paradeigmatikˆ dÔo eurèwc gnwstèc kai epituqhmènec arqitektonikèc Diakritik¸n Montèlwn me Bajiˆ Sunelikti-

kˆ Neurwnikˆ DÐktua, afenìc gia na tonÐsoume qarakthristikˆ touc ta opoÐa èqoun kajierwjeÐ kai se arqitektonikèc Baji¸n Paragwgik¸n Montèlwn kai afetèrou giatÐ kˆnoume qr sh amfotèrwn aut¸n twn montèlwn gia thn axiolìghsh twn eikìnwn pou èqoun paraqjeÐ apì GANs.

Diakritikˆ Montèla me Suneliktikˆ Neurwnikˆ DÐktua

Xekin¸ntac apì to LeNet-5 [11], ta Suneliktikˆ Neurwnikˆ DÐktua (CNN) èqoun lÐgo polÔ mia kajierwmènh dom : stoibagmènec suneliktikèc str¸seic akoloujoÔmenec apì str¸seic kanonikopoÐhshc kai omadopoÐhshc kai sto tèloc akoloujoÔmenec apì mÐa perissìterec pl rwc sundedemènec str¸seic. Parallagèc autoÔ tou basikoÔ sqediasmoÔ eÐnai arketˆ diadedomènec sth bibliografÐa gia taxinìmhsh eikìnwn kai èqoun shmei¸sei ta kalÔtera apotelèsmata mèqri s mera sta eurèwc gnwstˆ sÔnola dedomènwn eikìnwn, ìpwc sto MNIST [10], sto CIFAR [23] kai kurÐwc sto ImageNET [16] kai sthn prìklhsh taxinìmhshc eikìnwn ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [33], [78]. Skopìc thc teleutaÐac tan h ekpaÐdeush enìc montèlou se 1 ekatommÔrio eikìnec to opoÐo ja mporeÐ swstˆ na taxinom sei 100.000 ˆllec eikìnec se 1.000 tˆxeic. Gia ta megˆla sÔnola dedomènwn, ìpwc to ImageNET, h prìsfath tˆsh tan h aÔxhsh tou arijmoÔ kai tou megèjouc twn str¸sewn (enn. tou arijmoÔ suneliktik¸n fÐltrwn anˆ str¸sh), en¸ gÐnontan qr sh thc teqnik c Dropout [35] gia thn antimet¸pish tou over tting pou prokaloÔntan apì thn aÔxhsh thc qwrhtikìthtac (capacity) tou montèlou. Parakˆtw faÐnetai èna diˆgramma me thn exèlixh twn arqitektonik¸n Suneliktik¸n Neurwnik¸n DiktÔwn sto ImageNET, to opoÐo epibebai¸nei to parapˆnw. —lec autèc eÐnai arqitektonikèc me Suneliktikˆ Neurwnikˆ DÐktua. Sto sq ma faÐnetai h tˆsh gia aÔxhsh tou arijmoÔ twn str¸sewn, eidikˆ metˆ th qr sh twn residual connections to 2015 [41]. EpÐshc faÐnetai h apìtomh pt¸sh tou error rate (aristerìc kˆjetoc ˆxonac) to 2014 me thn parousÐash twn montèlwn VGGNet kai Inception pou analÔontai parakˆtw.

DÔo axioshmeÐwtec arqitektonikèc montèla gia taxinìmhsh eikìnwn me Suneliktikˆ Neurwnikˆ DÐktua, ta opoÐa efarmìzoume kai sth paroÔsa ergasÐa gia exagwg qarakthristik¸n (feature extraction) twn eikìnwn me skopì th sÔgkris touc, eÐnai ta VGGNet kai Inception. AkoloujeÐ mÐa sÔntomh perigraf twn montèlwn aut¸n estiˆzontac se stoiqeÐa thc arqitektonik c touc pou ta èqoun kajier¸sei wc ta pio dhmofil montèla gia taxinìmhsh eikìnwn kai exagwg qarakthristik¸n me Suneliktikˆ Neurwnikˆ

Sq ma 3: Nik triec arqitektonikèc me Suneliktikˆ Neurwnikˆ DÐktua thc prìklhshc taxinìmhshc eikìnwn apì to sÔnolo dedomènwn ImageNET.
Phg : CS231n: Convolutional Neural Networks for Visual Recognition, Fei-Fei Li, Standford University 2018 [94]

DÐktua.

Melèth perÐptwshc: VGGNet

To VGGNet[34] eÐnai èna montèlo gia taxinìmhsh eikìnwn qrhsimopoi¸ntac Suneliktikˆ Neurwnikˆ DÐktua, to opoÐo epino jhke to 2014 apì touc Simonyan kai Zisserman sto Visual Geometry Group tou PanepisthmÐou thc Oxfìrdhc. MazÐ me to GoogleNet pou analÔetai sth sunèqeia apotèlesan touc nikhtèc tou ImageNet Large-Scale Visual Recognition Challenge to 2014, petuqaÐnontac error rate 7.3% kai 6.7% antÐstoiqa sthn taxinìmhsh eikìnwn, shmei¸nontac shmantik meÐwsh se sqèsh me ton prohgoÔmeno nikht .

Basikˆ stoiqeÐa pou diaforopoioÔn thn arqitektonik tou VGGNetapì tic prohgoÔmenec eÐnai ta ex c:

- QrhsimopoÐhsh mikr¸n suneliktik¸n fÐltrwn, diˆstashc 3 Ö3. Autì epètreye thn aÔxhsh tou {bˆjouc} tou montèlou, allˆ kai thc shmˆntik c beltÐwshc thc poiìthtac twn exagìmenwn qarakthristik¸n stic endiˆmesec str¸seic. Autì ofeÐletai sto ìti trÐa (3) fÐltra diˆstashc 3 Ö3 (to èna metˆ to ˆllo) èqoun to Ðdio dektikì pedÐo (receptive eld) me èna fÐltro 7 Ö7, san autˆ pou qrhsimopoioÔntan sto

AlexNet [24] gia parˆdeigma, perièqontac ìmwc sqedìn tic misèc ekpaideÔsimec paramètrouc.

- Shmantikˆ megalÔteroc arijmìc suneliktik¸n str¸sewn. Stic dÔo parallagèc tou VGGNetpou qrhsimopoioÔntai eurèwc o arijmìc autìc eÐnai dekaèxi (16) gia to VGG16kai dekaennèa (19) gia to VGG19. Gia sÔgkrish sto montèlo AlexNet tou 2012 o arijmìc autìc tan mìlic oqt¸ (8).

Sto sq ma 4, parakˆtw, paratÐjentai sqhmatikˆ oi arqitektonikèc twn montèlwn VGG16 kai VGG19gia optikopoÐhsh thc dom c touc. To VGG16(aristerˆ) èqei 16 suneliktikèc str¸seic kai str¸seic omadopoÐhshc (pooling) prin tic fully-connected (FC) str¸seic, ìlec me fÐltra diastˆsewn 3 Ö3. AntÐstoiqa, to VGG19 (dexiˆ) èqei 19. EpÐshc, ìso kontinìtera proc thn èxodo briskìmaste, tìso auxˆnetai o arijmìc twn fÐltrwn kˆje suneliktik c str¸shc, en¸ mei¸netai twn m koc kai plˆtoc thc eisìdou thc str¸shc. Oi str¸seic me thn etikèta {PS} eÐnai pl rwc-sundedemènec (fully-connected) str¸seic, autèc me thn èndeixh {sunel} eÐnai suneliktikˆ fÐltra twn anagrafìmenwn diastˆsewn (aristerˆ thc etikètac) kai tou anagrafìmenou arijmoÔ (dexiˆ), en¸ oi str¸seic me thn etikèta {OmadopoÐhsh} efarmìzoun meÐwsh sto misu plˆtouc kai m kouc krat¸ntac to mègisto apì kˆje diakrit tetrˆda eikonostoiqeÐwn (max-pooling).

Melèth perÐptwshc: Inception

To montèlo Inception InceptionNet parousiˆsthke thn Ðdia qroniˆ me to VGGNet, dhlad sto ImageNet Large-Scale Visual Recognition Challenge tou 2014. Arqikˆ, eÐqe onomasteÐGoogleNet [36] kaj¸c oi eptˆ (7) apì touc ennèa (9) dhmiourgoÐ tou tan apì thn Google. H logik twn dhmiourg¸n tan na xefÔgoun apì thn tˆsh suneqoÔc aÔxhshc tou bˆjouc twn sunlektik¸n montèlwn kai na estiˆsoun se apodotikèc topikèc topologÐec diktÔwn. Çtsi, basikì domikì stoiqei¸n twn montèlwn tÔpou Inception eÐnai èna upodÐktuo apoteloÔmeno apì parˆllhlec suneliktikèc str¸seic (me diaforetikˆ megèjh fÐltrwn) kai str¸seic omadopoÐhshc, to Inception Module, to opoÐo faÐnetai sto sq ma 5 pou akolouje. Prìkeitai gia mÐa network-in-network topologÐa pou epi-trèpei se kˆje module na exˆgei akìma kalÔtera kai pio akrib qarakthristikˆ twn eikìnwn eisìdou. Sthn apl (naive) ekdoq touc, ta Inception Modules kˆnoun parˆllhla suneliktikˆ filtrarÐsmata me fÐltra diastˆsewn 1 Ö1, 3Ö3 kai 5Ö5, en¸ sthn kanonik ekdoq touc (pou qrhsimopoi jhke kai sto GoogleNet) gÐnetai meÐwsh twn kanali¸n (channels) thc eisìdou me ekpaideÔsima suneliktikˆ fÐltra 1 Ö1 prin aut {eisèljei}

(a) VGG16

(b) VGG19

Sq ma 4: Arqitektonikèc twn montèlwn VGG16 kai VGG19 pou anadeÐqjhkan nik triec sth prìklhsh entopismoÔ kai taxinìmhshc, antÐstoiqa, eikìnwn apì to sÔnolo dedomènwn ImageNET to 2014.
Phg : Anakataskeu apì Very Deep Convolutional Networks for Large-Scale Image Recognition , Karen Simonyan and Andrew Zisserman, 2014[34]

sta suneliktikˆ fÐltra diastˆsewn 3 Ö3 kai 5 Ö5. Autì eÐqe wc apotèlesma shmantik meÐwsh tou upologistikoÔ kìstouc tou kˆje emprìsjiou perˆsmatoc (forward pass) tou montèlou kai èkane dunat thn ekpaÐdeus tou.

(a) Aplì (Naive) Inception Module

(b) Inception Module me MeÐwsh Kanali¸n

Sq ma 5: Inception Modules: Basikˆ domikˆ stoiqeÐa tou Inception.
Phg : Anakataskeu apì Going Deeper with Convolutions , Szegedy et al., 2014 [36]

Autì pou epitugqˆnetai me th qrhsimopoÐhsh twn Inception Modules eÐnai ìti to kˆje èna apì autˆ {blèpei} thn èxodo tou prohgoÔmenou me diaforetikˆ dektikˆ pedÐa (receptive elds) , me apotèlesma oi prˆxeic eswterikˆ tou kajenìc na gÐnontai me megalÔterh apodotikìthta.

Mia ˆllh shmantik kainotomÐa tou GoogleNet kai twn montèlwn tÔpou Inception eÐnai h mh-qrhsimopoÐhsh pl rwc-sundedemènwn str¸sewn prin thn èxodo, plhn mÐac megèjouc Ðdiou me tou arijmoÔ twn tˆxewn exìdou prokeimènou na mporeÐ na efarmosteÐ h str¸sh exagwg c pijanot twn, Softmax [17]. Basizìmenoi sthn parat rhsh ìti h suntriptik pleioyhfÐa twn paramètrwn entopÐzontai stic pl rwc-sundedemènec str¸seic prin thn èxodo antÐstoiqwn montèlwn, oi dhmiourgoÐ tou GoogleNet afaÐresan autèc tic str¸seic, kˆti pou touc epètreye na ftiˆxoun arketˆ pio sÔnjetec domèc exagwg c qarakthristik¸n, dhlad ta Inception Modules.

Tèloc, oi dhmiourgoÐ tou GoogleNet èqoun prosjèsei dÔo akìma {bohjhtikoÔc} taxi-nomhtèc (touc {taxinomht c 0} kai {taxinomht c 1} pou faÐnontai sto sq ma 6). Oi bohjhtikoÐ taxinomhtèc èqoun Ðdia morf exìdou me ton basikì taxinomht , qrhsimopoi-jhkan prokeimènou na stajeropoi soun kai na kˆnoun pio apodotik thn ekpaÐdeush tou diktÔou, en¸ den qrhsimopoioÔntai katˆ th fˆsh dokim c (evaluation). Sto sq ma 6 parakˆtw, paratÐjetai sqhmatikˆ h basik dom tou GoogleNet. EkeÐ, faÐnontai ta stoibagmèna (stacked) Inception Modules kaj¸c kai oi bohjhtikoÐ taxinomhtèc. Aut eÐnai mia perigraf se uyhlì epÐpedo. Gia pio leptomer perigraf o anagn¸sthc kaleÐtai na sumbouleuteÐ to ˆrjro pou paratÐjetai sthn phg tou sq matoc.

H arqitektonik tou GoogleNet (Inception v1) pèrase apì mia seirˆ epanalhptik¸n beltiÛsewn apì touc dhmiourgoÔc tou. Çtsi, stadiakˆ parousiˆsthkan ta montèla Inception v2 kai Inception v3 [46], me èmfash sthn aÔxhsh thc apìdoshc sto eswterikì tou kˆje Inception Module kai thc axiopistÐac twn bohjhtik¸n taxinomht¸n. EpÐshc, to 2016, parousiˆsthkan ta Inception v4 kai Inception-ResNet [60], me peraitèrw beltiÛseic thc apodotikìthtac twn Inception Modules kaj¸c, eisagwg residual connections (Inception-ResNet v1)kai hyperparameter tuning (Inception-ResNet v2).

Bajiˆ Paragwgikˆ Montèla

Epanerqìmenoi sthn Paragwgik MontelopoÐhsh, metˆ th sÔntomh antiparabol twn Diakritik¸n Montèlwn kai thn parˆjesh dÔo (2) eurèwc diadedomènwn arqitektonik¸n aut¸n me DÐktua Bajiˆc Mˆjhshc, sthn paroÔsa upoenìthta ja gÐnei mia eisagwg sta Paragwgikˆ Montèla me DÐktua Bajiˆc Mˆjhshc Bajiˆ Paragwgikˆ Montèla (BPM) (Deep Generative Models). Ta BPM an koun sta montèla Statistik c Paragwgik c MontelopoÐhshc kai ousiastikˆ eÐnai Teqnhtˆ Neurwnikˆ DÐktua (TND) me mÐa perissìterec (sun jwc pollèc) krufèc str¸seic, ta opoÐa ekpaideÔontai qrhsimopoi¸ntac megˆla sÔnola dedomènwn prokeimènou na ektim soun na proseggÐsoun sÔnjetec katanomèc pijanìthtac me uyhl diastasimìthta. Çna epituq¸c ekpaideumèno BPM qrhsimopoieÐtai sth sunèqeia eÐte gia paragwg nèwn shmeÐwn (dhl. dedomènwn) apì to q¸ro eisìdou, ta opoÐa akoloujoÔn thn Ðdia katanom pijanìthtac me aut twn dedomènwn ekpaÐdeushc, gia thn ektÐmhsh thc pijanofˆneiac (likelihood) thc kˆje parat rhshc.

Oi efarmogèc twn BPM eÐnai pollèc kai poikÐlec, me kˆpoiec ìpwc ta legìmena Deep Fakes [105], [110] na èqoun proselkÔsei dhmosiìthta lìgw tou kindÔnou pou elloqeÔei o realismìc twn paragìmenwn teqnht¸n eikìnwn kai bÐnteo. Seirˆ mejìdwn emfanÐzontai sth bibliografÐa, wstìso, kai se episthmonikoÔc tomeÐc efarmog¸n, ìpou h mhqanik mˆjhsh jewroÔntan paradosiakˆ anÐkanh na eisèljei, ìpwc h aÔxhsh thc anˆlushc eikìnwn (Image Super-Resolution) me GANs [54], h deigmatolhyÐa katastˆsewn tou ShmeÐou IsorropÐac se sust mata poll¸n swmˆtwn [106] sth fusik kai h dhmiourgÐa parametrik¸n sunjetik¸n prosomoi¸sewn [92] sth mhqanik reust¸n.

›swc thn pio epanastatik efarmog BPM èwc s mera apoteleÐ h prìsfath parousÐash apì thn NVIDIA enìc sust matoc pou qrhsimopoieÐ GANs gia sumpÐesh bÐnteo se peribˆllon zwntan c metˆdoshc, to opoÐo paratÐjetai sto sq ma 7 parakˆtw. Prìkeitai

Sq ma 6: Montèlo GoogleNet Inception v1 ìpwc alli¸c onomˆsthke.

Phg : Anakataskeu apì Going deeper with convolutions , Szegedy et al. , 2014 [36]

gia ton AI Video Codec tou sust matoc NVIDIA Maxine®, ìpou o pompìc stèlnei to arqikì frame wc eikìna kai gia kˆje epìmeno stèlnei tic jèseic 126 shmeÐwn (keypoints) sta ìria twn qarakthristik¸n tou pros¸pou tou omiloÔnta. O dèkthc lambˆnei thn arqik eikìna kai gia kˆje epìmenh lambˆnei ta keypoints ta opoÐa ta metasqhmatÐzei se disdiˆstatec eikìnec. Katìpin, sth meriˆ tou dèkth qrhsimopoieÐtai èna pro-ekpaideumèno GAN gia metasqhmatismì thc eikìnac twn keypoints, me sunj kh thn arqik eikìna, sto trèqon frame tou bÐnteo. —pwc analÔetai sto ˆrjro sthn istoselÐda thc etaireÐac, autì odhgeÐ se drastik meÐwsh tou apaitoÔmenou eÔrouc z¸nhc metˆdoshc dedomènwn gia Ðdia telik poiìthta bÐnteo, sthn anabˆjmish thc poiìthtac tou paragìmenou bÐnteo gia Ðdio rujmì metˆdoshc kai sth dunatìthta dhmiourgÐac pleiˆdac teqnht¸n efè bÐnteo apl¸c allˆzontac thn arqik eikìna - sunj kh.

Sq ma 7: AI Video Codec tou sust matoc Maxine® thc NVIDIA.
Phg : NVIDIA MAXINE: Accelerated SDK with state-of-the-art AI features for building virtual collaboration and content creation applications , NVIDIA (https://developer.nvidia.com/maxine)

Stic enìthtec pou katalambˆnoun to upìloipo autoÔ tou kefalaÐou, ja gÐnei mia parousÐash diafìrwn tÔpwn BPM pou èqoun katˆ kairoÔc parousiasteÐ sth bibliografÐa kai èqoun efarmosteÐ sth prˆxh. Çtsi, ja anafèroume touc treic plèon eurèwc qrhsimopoioÔmenouc tÔpouc BPM: ta AutopalindronoÔmena Paragwgikˆ Montèla, touc Autìmatouc Kwdikopoihtèc kai fusikˆ twn Generative Adversarial Networks pou apoteloÔn kai ton pur na thc paroÔsac ergasÐac. Gia kˆje ènan apì autoÔc, ja gÐnei parˆjesh sqetik¸n montèlwn kai teqnik¸n pou anaptuqjeÐ kaj¸c kai apotelesmˆtwn apì th qr sh touc.

2.2 AutopalindronoÔmena Bajiˆ Paragwgikˆ Montèla

Ta AutopalindronoÔmena (autoregressive) Paragwgikˆ Montèla apoteloÔn mÐa apì tic pr¸tec teqnikèc gia Paragwgik MontelopoÐhsh. H basik idèa pou kinhtrodìthse th dhmiourgÐa aut¸n twn mejìdwn eÐnai o kanònac thc alusÐdac sth JewrÐa Pijanot twn:

$$p(\mathbf{x}) = p(x_1)\ \ p(x_2|x_1)\ \ \cdots\ \ p(x_n|x_1, \ldots, x_{n-1})$$

ìpou gia th prosèggish twn desmeumènwn pijanot twn èqoun protajeÐ poÐkillec teqnikèc, me kÔriec th qr sh Palindrìmhshc pou eÐqe arqikˆ dokimasteÐ [20], [39]) kai th qr sh Baji¸n Neurwnik¸n DÐktuwn pou dokimˆsthkan argìtera me megˆlh epituqÐa, ìpwc analÔetai kai se epìmenh upoenìthta.

Sth perÐptwsh Paragwgik c MontelopoÐhshc eikìnwn, h eikìna jewreÐtai wc mia diatetagmènh akoloujÐa, $\mathbf{x}$, tuqaÐwn metablht¸n (TM), ìpou h kˆje TM antiproswpeÔei th pijanìthta to antÐstoiqo eikonostoiqeÐo na pˆrei tim 1/0 (asprìmaurec eikònec - duadikèc TM) mÐa apì tic 256 timèc (0=maÔro èwc 255=ˆspro gia eikònec se diabajmÐseic tou gkri). H diˆtaxh twn TM arqikˆ gÐnontan ìpwc kai se èna sarwt eikònwn (raster scan order), dhlad gramm -gramm kai apì aristerˆ proc dexiˆ se kˆje gramm . Çtsi, kai sÔmfwna me thn parapˆnw exÐswsh, x_1 ja eÐnai to epˆnw-aristerˆ eikonostoiqeÐo miac eikònac, x_n to kˆtw dexiˆ kai to n ja isoÔtai me W H (W: plˆtoc eikònac se eikonostoiqeÐa, H: Ôyoc eikònac - arqikˆ jewroÔntan mìno asprìmaurec eikònec kai eikònec se diabajmÐseic tou gkri, ˆra o arijmìc twn kanali¸n eÐnai èna).

Oi teqnikèc Paragwgik c MontelopoÐhshc eikìnwn me AutopalindronoÔmena Paragwgikˆ Montèla mporoÔn na qwrisjoÔn se duo basikèc oikogèneiec, autèc stic opoÐec gia thn prosèggish katanom¸n qrhsimopoieÐtai h Logistik Palindrìmhsh (Logistic Regression) kai se autèc stic opoÐec qrhsimopoioÔntai Epanalambanìmena Neurwnikˆ DÐktua (Recurrent Neural Networks) gia ton Ðdio skopì. AkoloujeÐ mia sÔntomh episkìphsh amfotèrwn aut¸n twn oikogenei¸n me parˆjesh antiproswpeutik¸n paradeigmˆtwn apì th bibliografÐa.

MontelopoÐhsh basismènh sth Logistik Palindrìmhsh

To montèlo Fully-Visible Sigmoid Belief Network (FVSBN) [39], gia parˆdeigma, qrhsimopoioÔse Logistic Regression gia upologismì twn upo-sunj kh pijanot twn se asprìmaurec eikònec me th seirˆ diˆtaxhc twn eikonostoiqeÐwn ìpwc tou skˆner. Çtsi, o kˆje

ìroc tou ginomènou tou kanìna alusÐdac (plhn twn pr¸twn dÔo pou eÐnai tetrimmènec peript¸seic) ja eÐnai:

$$p(X_i = 1 | x_1, :::, x_{i-1}; \sim^i) = \left({}^i_0 + \sum_{j=1}^{i-1} {}^i_j x_j \right)$$

ìpou to diˆnusma $\sim^i = [{}_0; :::; {}_{i-1}]$ eÐnai oi(i) parˆmetroi thc Logistic Regression kai $fx\ g$ eÐnai oi timèc pou èlaban oi prohgoÔmenec (i 1) TM. Epeid prìkeitai gia asprìmaurec eikìnec, sthn ousÐa h èxodoc tou ajroÐsmatoc thc Logistik c Palindrìmhshc apoteleÐ ektÐmhsh thc paramètrou p thc Bernoulli katanom c pou akolouje h ekˆstote TM. AntÐstoiqa, to montèlo Neural Autoregressive Distribution Estimator (NADE) [20] qrhsimopoieÐ èna TND kai en suneqeÐa mLogistic Regression antÐ gia mìno thn teleutaÐa gia thn ektÐmhsh thc kˆje desmeumènhc pijanìthtac. To kèrdoc apì th qr sh tou TND eÐnai h ekmˆjhsh miac mh grammik c sqèshc twn metablht¸n eisìdou kai to pèrasma tou apotelèsmatoc aut c sth palindrìmhsh.

—pwc faÐnetai kai sto sq ma 8, o upologismìc thc pijanofˆneiac nèwn eikìnwn me paragwgikˆ montèla tÔpou FVSBN NADExekinˆei apì to pr¸to aristerˆ eikonostoiqeÐo kai sunqÐzei proc to kˆtw dexiˆ, me kˆje epìmeno eikonostoiqeÐo na qrhsimopoieÐ san eÐsodo sto antÐstoiqo TND tic timèc ìlwn twn prohgoÔmenwn eikonostoiqeÐwn thc eikìnac. Oi sundèseic me mple sto montèlo NADE tou sq matoc dhl¸noun diamoirasmì twn bar¸n pou antistoiqÐzontai se kˆje parat rhsh prin thn eÐsodì thc sta TND. Gia ton telikì upologismì thc $p(\forall)$, an $p.q.\forall = [1; 0; 0; 1]$, tìte $p(\forall) = \hat{v}_1 \ (1 \ \hat{v}_2) \ (1 \ \hat{v}_3) \ \hat{v}_4$, afoÔ to $\hat{v}_i = p(V_i = 1 | v_1, :::, v_{i-1}; \sim^i)$ (ìpou sth perÐptwsh tou NADE oi parˆmetroi $\sim^i$ perilambˆnoun kai tic paramètrouc tou TND ektìc apì autèc thc Logistic Regression). Sto sq ma 9 faÐnontai paragwgèc eikìnwn apì to NADE ekpaideumèno sto sÔnolo dedomènwnMNIST digits. Autì pou eikonÐzetai sta dexiˆ eÐnai deÐgmata apì ta $\hat{v}_i$'s gia $i = 1; :::; 28^2$ (sthn ousÐa eÐnai oi parˆmetroi thc katanom c Bernoulli). Tèloc, eÐnai skìpimo na anaferjeÐ ìti mÐa beltÐwsh tou teleutaÐou, apoteleÐ to montèlo Masked Autoregressive Density Estimator (MADE) [40], to opoÐo qrhsimopoieÐ mia parallag enìc autìmatou kwdikopoiht gia to metasqhmatismì twn metablht¸n eisìdou kai mˆskec bar¸n gia na exasfalisteÐ ìti kˆje èxodoc ephreˆzetai mìno apì tic prohgoÔmenec eisìdouc, kai ˆra ìti o autìmatoc kwdikopoiht c leitourgeÐ wc èna autopalindronoÔmeno paragwgikì montèlo, arketˆ ìmwc pio apodotikˆ (parˆllhloc upologismìc twn exìdwn).

Sq ma 8: Apeikìnish thc diadikasÐac upologismoÔ pijanofˆneiac (antÐstoiqh qrhsimopoieÐtai kai gia paragwg) Bernoulli akoloujÐac me tèssera (4) stoiqeÐa apì montèla tÔpou FVSBN(aristerˆ) kai NADE(dexiˆ).
Phg : CS236: Deep Generative Models , Standford (https://deepgenerativemodels.github.io)

Sq ma 9: DeÐgmata eikìnwn (dexiˆ) pou èqoun paraqjeÐ apì to autopalindronoÔmeno montèlo NADEto opoÐo ekpaideÔthke se qeirìgrafa yhfÐa apì to sÔnolo dedomènwn tou MNIST (aristerˆ).
Phg : The Neural Autoregressive Distribution Estimator , Larochelle et. al , 2015 [20]

MontelopoÐhsh me Epanalambanìmena Neurwnikˆ DÐktua

Orm¸menoi apì thn idèa ìti ta Epanalambanìmena Neurwnikˆ DÐktua (ENN) èqoun sqediasteÐ gia kai qrhsimopoihjeÐ eurèwc se akoloujiakèc eisìdouc me {makrèc} exart seic metaxÔ twn deigmˆtwn touc, polloÐ ereunhtèc strˆfhkan se autˆ prokeimènou na montelopoi soun thc desmeumènec katanomèc se efarmogèc Paragwgik c MontelopoÐhshc megˆlwn akolouji¸n, ìpwc eÐnai mia eikìna èna hqhtikì s ma. Gia mÐa megˆlh akoloujÐa eisìdou me makrinèc exart seic, den eÐnai upologistikˆ efiktì (tractable) na sqediasteÐ montèlo pou {jumˆtai} ìlec tic prohgoÔmenec eisìdouc. Gia to lìgo autì

ta ENN diathroÔn mia eswterik katˆstash (sthn opoÐa jewrhtikˆ èqoun sugkrathjeÐ basikˆ qarakthristikˆ twn deigmˆtwn) thn opoÐa kai anane¸noun kaj¸c dèqontai ìlo kai neìtera deÐgmata sthn eÐsodo.

Mia apl morf enìc ENN faÐnetai sto sq ma 10 parakˆtw. H eswterik katˆstash, h_t, diathreÐ mÐa {perÐlhyh} twn eisìdwn èwc th qronik stigm t, en¸ h èxodoc gia th perÐptwsh twn APM qrhsimopoieÐtai gia na parametropoi sei kˆpoia upo-sunj kh katanom , $p(x_t|x_{1:t\ 1})$. To dÐktuo ekpaideÔetai gia na beltistopoi sei tic timèc twn paramètrwn tou, W_{xh}, W_{hh} kai W_{ho}, ¸ste mèsw aut¸n to montèlo na anajètei th mègisth pijanofˆneia sta dedomèna ekpaÐdeushc. Çtsi, me thn pˆrodo tou qrìnou to dÐktuo metabˆllei thn eswterik tou katˆstash kaj¸c kai ta bˆrh me ta opoÐa upologÐzontai oi èxodoi apì thn katˆstash aut . Oi èxodoi me th seirˆ touc, gia thn perÐptwsh Paragwgik¸n Montèlwn, mporoÔn na jewrhjoÔn ìti proseggÐzoun tic desmeumènec katanomèc pijanìthtac, $p(x_t|x_{1:t\ 1})$.

Sq ma 10: Aplì ENN mÐac eisìdou kai mÐac exìdou (faÐnetai kai {xediplwmèno} gia na fanoÔn ta qronikˆ b mata pou apaitoÔntai gia ton upologismì twn exìdwn).
Phg : Anakataskeu apì Wikimedia Commons: A diagram for a one-unit recurrent neural network , fdeloche, 2017

Sta plaÐsia Paragwgik c MontelopoÐhshc eikìnwn me AutopalindronoÔmena ENN, Ðswc h pio antiproswpeutik doulieˆ pou èqei parousiasteÐ sth bibliografÐa eÐnai to montèlo PixelRNN tou van den Oord to 2016 [63]. EkeÐ, gia thn paragwg eikìnwn allˆ kai thn anˆjesh pijanofˆneiac se autèc qrhsimopoioÔntan h seirˆ tou sarwt , en¸ to montèlo sqediˆsthke gia èggrwmec eikìnec (RGB). Autì shmaÐnei pwc h desmeumènh katanom gia kˆje eikonostoiqeÐo apaiteÐ ton orismì tri¸n qrwmˆtwn, o opoÐoc epi-lèqjhke na gÐnei epÐshc akoloujiakˆ (pr¸ta to kìkkino, metˆ prˆsino kai tèloc to

mple):

$$p(x_t j x_{1:t\ 1}) = p(x_t^{red} j x_{1:t\ 1})\quad p(x_t^{green} j x_{1:t\ 1}; x_t^{red})\quad p(x_t^{blue} j x_{1:t\ 1}; x_t^{red}; x_t^{green})$$

en¸h kˆje epimèrouc desmeumènh katanom eÐnai kathgorik (multinoulli) me 256 pijanèc timèc.

Mia beltÐwsh tou parapˆnw basizìmenh sthn antÐlhyh ìti h tim enìc eikonostoiqeÐou ephreˆzetai prwtÐstwc apì tic timèc twn geitonik¸n eikonostoiqeÐwn pou èqoun dh paraqjeÐ, tan to AutopalindronoÔmeno Paragwgikì Montèlo PixelCNN [64] sto opoÐo qrhsimopoi jhkan Suneliktikèc Str¸seic antÐ gia BND. Apotèlesma autoÔ, tan h polÔ taqÔterh kai pio apotelesmatik ekpaÐdeush tou montèlou gia Ðdiac {poiìthtac} paragìmena apotelèsmata se sqèsh me ton prokˆtoqì tou. Sto sq ma 11 parousiˆzetai grafikˆ h diadikasÐa paragwg c eikìnwn apì autˆ ta montèla.

Tèloc, gia lìgouc plhrìthtac, skìpimo eÐnai na anaferjeÐ ìti parìmoiac logik c AutopalindronoÔmena Paragwgikˆ Montèla me ENN èqoun efarmosteÐ me epituqÐa kai se efarmog c Paragwgik c MontelopoÐhshc qou. To montèlo WaveNet [62] apoteleÐ èwc s mera Ðswc to pio apotelesmatikì montèlo gia metatrop keimènou se omilÐa (Text-to-Speech) paragwg mousik c (Music Generation), en¸ upˆrqei apì to 2018 stic perissìterec kinhtèc suskeuèc me leitourgikì Android.

(a) Paragwg eikìnac pixel-by-pixel (PixelRNN)

(b) Qr sh mask¸n se dÐktua tÔpou LSTM

(g) Efarmog sunelikti-k¸n diktÔwn gia taqÔthta (PixelCNN)

(d) Qr sh ma-sk¸n sta sune-liktikˆ fÐltra

Sq ma 11: Apeikìnish thc diadikasÐac paragwg c eikìnwn apì paragwgikˆ montèla me ENN, PixelRNN (aristerˆ) kai PixelCNN (dexiˆ).

Phg : Pixel Recurrent Neural Networks kai Conditional Image Generation with PixelCNN Decoders , van den Oord et al., 2016 [63], [64]

2.3 Paragwgik MontelopoÐhsh me Autìmatouc Kwdikopoihtèc

Sth paroÔsa upoenìthta ja parousiˆsoume touc Autìmatouc Kwdikopoihtèc (AK) kai kurÐwc mia eidik ekdoq touc, touc MetablhtoÔc Autìmatouc Kwdikopoihtèc (MAK) (Variational Autoencoders - VAEs), oi opoÐoi qrhsimopoioÔntai suqnˆ gia Paragwgik MontelopoÐhsh. Aut h parousÐash, ìpwc anafèrjhke kai prohgoÔmena, gÐnetai diìti pr¸ton qrhsimopoioÔntai suqnˆ parallagèc twn MAK wc domikˆ stoiqeÐa Paragwgik¸n Montèlwn me PAD kai deÔteron giatÐ oi MAK apotelèsan tic pr¸tec prospˆjeiec Paragwgik c MontelopoÐhshc uyhl c-anˆlushc eikìnwn me axiìloga apotelèsmata. Epiprìsjeta, oi AK eÐnai mÐa apì tic teqnikèc kai idèec pou apoteloÔn antikeÐmeno èreunac ed¸ kai dekatÐec sto tomèa thc Mhqanik c Mˆjhshc me TND [5], [6], [8].

Autìmatoi Kwdikopoihtèc

—pwc perigrˆfetai kai sto ìnomˆ touc, oi Autìmatoi Kwdikopoihtèc eÐnai ènac tÔpoc TND pou mac bohjˆei na kwdikopoi soume dedomèna {autìmata}, dhl. na broÔme apodotikoÔc k¸dikec gia èna sÔnolo dedomènwn qwrÐc epÐbleyh (unsupervised learning). —pwc faÐnetai kai sto sq ma 12 parakˆtw, apoteloÔntai apì dÔo mèrh: ton kwdikopoiht (encoder) pou {majaÐnei} mÐa sunˆrthsh $\vec{z} = f(\vec{x})$ gia na metasqhmatÐzei thn èisodo $\vec{x}$ se èna diˆnusma $\vec{z}$, pou onomˆzetai kai endiˆmesh anaparˆstash kai ton apokwdikopoiht (decoder) pou {majaÐnei} thn antÐstrofh sunˆrthsh $\vec{x}' = g(\vec{z}) = g(f(\vec{x}))$ gia na anakataskeuˆsei thn arqik eÐsodo. Çtsi, h eikìna eisìdou arqikˆ pernˆei apì ton Kwdikopoiht (encoder) kai metasqhmatÐzetai se èna diˆnusma ston lanjˆnonta q¸ro shmantikˆ mikrìterhc diˆstashc. AkoloÔjwc, o Apokwdikopoiht c (decoder) lambˆnei to diˆnusma autì kai to metasqhmatÐzei se mÐa eikìna Ðdiac diˆstashc me thn arqik . Tèloc, oi dÔo eikìnec sugkrÐnontai, prokÔptei to kìstoc kai oi parˆmetroi twn diktÔwn anane¸ntai me skopì th meÐwsh autoÔ sÔmfwna me kˆpoion algìrijmo beltistopoÐhshc.

O skopìc kai h kÔria qr sh twn AK eÐnai h ekmˆjhsh aut c thc kwdikopoÐhshc prokeimènou na {sumpièsoume} apodotikˆ ta dedomèna mei¸ntac th diastasimìthtˆ touc (dimensionality reduction) na mˆjoume basikˆ touc qarakthristikˆ (feature learning). Oi AK ekpaideÔontai mei¸ntac to sfˆlma anakataskeu c (reconstruction loss) , en¸ gia th beltÐwsh twn paramètrwn touc sun jwc qrhsimopoieÐtai o algrìrijmoc back-propagation gia eÔresh thc kateÔjunshc belti¸shc se sunduasmì me kˆpoio algrìrijmo

beltistopoÐhshc (ìpwc oi Stochastic Gradient Descent, RMSProp k.a.).

Sq ma 12: Sqhmatik anaparˆstash thc leitourgÐac enìc Autìmatou Kwdikopoiht (AK).
Phg : Anakataskeu apì GANs in action: deep learning with generative adversarial networks , Langr et al., 2019[104]

Pollèc parallagèc twn apl¸n AK pou paraousiˆsthkan prohgoumènwc èqoun emfanisteÐ sth bibliografeÐa. Shmantikìterec ìmwc apì autèc kai pio diadedomènec eÐnai oi Kanonismènoi Autìmatoi Kwdikopoihtèc (KAK) (Regularized Autoencoders - RAEs)kai oi MetablhtoÐ Autìmatoi Kwdikopoihtèc (MAK). AkoloÔjwc, analÔetai h kˆje mÐa apì autèc tic parallagèc me antÐstoiqa paradeÐgmata montèlwn kai epituqhmènwn arqitektonik¸n, en¸ sto tèloc aut c thc upoenìthtac gÐnetai parˆjesh paradeigmˆtwn qr shc MAK se Paragwgik MontelopoÐhsh eikìnwn.

Kanonismènoi Autìmatoi Kwdikopoihtèc (KAK)

Gia thn eustaj ekpaÐdeush montèlwn me AK (idiaÐtera stic peript¸seic ìpou h diˆstash tou lanjˆnontoc dianÔsmatoc ≠ epitrèpetai na eÐnai megalÔterh apì aut thc eisìdou) ja prèpei h qwrhtikìthtec twn dÔo epimèrouc diktÔwn (tou kwdikopoiht kai apokwdikopoiht antÐstoiqa) kaj¸c kai h diˆstash tou dianÔsmatoc ≠ na epilègontai me bˆsh th poluplokìthta thc katanom c twn dedomènwn ekpaÐdeushc. Montèla me megˆlh qwrhtikìthta (dhl. dÐktua me pollèc str¸seic kai megˆlo arijmì ekpaideÔsimwn paramètrwn) ìtan kaloÔntai na mˆjoun mÐa sunˆrthsh anaktaskeu c gia sqetikˆ aplì

mikrì sÔnolo dedomènwn, mporeÐ eÔkola na odhghjoÔn sthn prosèggish thc $g(f(x))$ me th monadiaÐa sunˆrthsh.

Me aform autˆ ta eur mata, pollèc ereunhtikèc ergasÐec odhg jhkan sth qr sh Kanonismènwn (Regularized) AK prokeimènou na xefÔgoun apì aplˆ montèla TND me lÐgec str¸seic kai na mporèsoun na ekpaideÔsoun epituq¸c bajiˆ TND kai montèla me megˆlhc diastshc lanjˆnontec q¸rouc, qwrÐc autˆ apl¸c na majaÐnoun monadiaÐec sunart seic. Autì pou ousiastikˆ diaforopoieÐ touc KAK apì touc aploÔc AK eÐnai ìti plèon sth sunathsh kìstouc ektìc apì ton ìro pou metrˆ to sfˆlma anakataskeu c prostÐjentai ènac perissìteroi ìroi pou enjarrÔnoun to montèlo na èqei ki ˆllec qr simec idiìthtec ektìc thc metaforˆc thc eisìdou sthn exìdo. Tètoiec idiìthtec eÐnai gia parˆdeigma h endiˆmesh anaparˆstash na èqei arai (sparse) morf , to montèlo na mporeÐ na mei¸nei to jìrubo (denoising) thc eisìdou na anèqetai metabolèc aut c {sustèllontac} (contracting) tic paramètrouc tou. Oi pio diadedomènoi tÔpoi KAK analÔontai stic paragrˆfouc pou akoloujoÔn.

AraioÐ Autìmatoi Kwdikopoihtèc

Oi AraioÐ Autìmatoi Kwdikopoihtèc (Sparse Autoencoders) [14], [15] eÐnai Ðdioi me touc aploÔc AK me th prosj kh enìc ìrou sth sunˆrthsh kìstouc. Çtsi, h sunˆrthsh kìsouc den perièqei plèon mìno to sfˆlma anakataskeu c, $L(x; g(f(x)))$, allˆ kai mÐa poin {araiìthtac} thc endiˆmeshc anaparˆstashc z, (z). Epomènwc, to kìstoc gia kˆje eÐsodo stouc AraioÔc Autìmatouc Kwdikopoihtèc ja eÐnai:

$$L(x; f_{\tilde{}}; g_{\tilde{}}) = \text{ReconLoss}(x; g(f(x; f_{\tilde{}}); g_{\tilde{}})) + \quad (z = f(x; f_{\tilde{}}))$$

ìpou $f_{\tilde{}}$, $g_{\tilde{}}$ eÐnai oi ekpaideÔsimec parˆmetroi tou diktÔou kwdikopoi shc kai apokw-dikopoÐhshc antÐstoiqa, ReconLoss eÐnai mia sunˆrthsh apìstashc metaxÔ thc eisìdou kai thc anakataskeu c thc (p.q. L1/ Manhattan, L2/EukleÐdia k.a.) kai o suntelest c barÔthtac tou deÔterou ìrou. Sth perÐptwsh pou h sunˆrthsh poin c thc araiìthtac eÐnai hL1 (Manhattan), tìte o deÔteroc ìroc thc parapˆnw exÐswshc gÐnetai:

$$(z) = \quad \sum_i |z_i|$$

H parapˆnw sunˆrthsh apìstashc enjarÔnei to montèlo na anajètei mikrèc timèc sta stoiqeÐa tou endiˆmesou dianÔsmatoc. O Glorot et al. qrhsimopoÐhsan Rectified Linear Units (ReLUs) wc sunart seic energopoÐhshc se montèla Arai¸n Autìmatwn Kwdikopoiht¸n

[19] me apotelèsma na anatÐjontai akrib¸c mhdenikèc timèc se stoiqeÐa tou $\tilde{z}$ gia di^forec eisìdouc.

Basikèc efarmogèc twn Arai¸n AK eÐnai h arai kwdikopoÐhsh (sparse coding) [82] kaj¸c kai h exagwg qarakthristik¸n apì to sÔnolo dedomènwn [28] ta opoÐa mporoÔn na qrhsimopoihjoÔn en suneqeÐa apì k^poio algrìrijmo taxinìmhsh anagn¸rishc anwmali¸n (anomaly detection). Peraitèrw majhmatik an^lush thc ekpaÐdeushc Arai¸n AK èqei deÐxei pwc aut eÐnai isodÔnamh me thn ekpaÐdeush mègisthc-pijanof^neiac (maximum-likelihood) enìc paragwgikoÔ montèlou pou èqei lanj^nousec tuqaÐec metablhtèc prokeimènou na m^jei thn aposteriori sun^rthsh katanom c pijanìthtac, $p_{model}(\tilde{x}|\tilde{z})$ [70].

Denoising Autìmatoi Kwdikopoihtèc

Se antÐjesh me touc AraioÔc AK pou prosjètoun èna deÔtero ìro sth sun^rthsh kìstouc, oi Denoising apl¸c all^zoun ton trìpo upologismìu tou sf^lmatoc anakataskeu c prokeimènou na {m^joun} kalèc anaparast^seic. Se aut thn parallag twn apl¸n AK, loipìn, o kwdikopoiht c den dèqetai thn eikìna wc proc thn opoÐa upologÐzetai to sf^lma anakataskeu c, all^ mia enjìrubh edkoq thc (eikìna + jìruboc). Çtsi, ìpwc apeikonÐzetai kai sto sq ma 13 parak^tw, to dÐktuo kaleÐtai na apojorubopoi sei thn eikìna eisìdou prokeimènou na meiwjeÐ to sf^lma anakataskeu c wc proc thn arqik (qwrÐc jìrubo eikìna). Epomènwc, h nèa sun^rthsh kìstouc thn opoÐa ekpaideÔetai gia na elaqistopoi sei ènac Denoising AK eÐnai:

$$L(\tilde{x}; f^{\tilde{}}_f; {}^{\tilde{}}_g) = \text{ReconLoss}(\tilde{x}; g(f(\tilde{x} + \text{noise}; {}^{\tilde{}}_f); {}^{\tilde{}}_g))$$

SÔmfwna me ta parap^nw, oi Denoising AK prèpei na anairèsoun th di^brwsh thc eikìnac apì to jìrubo kai ìqi apl¸c na metafèroun thn eÐsodo sthn èxodo. H ekpaÐdeush AK me autìn ton trìpo pièzei tic f kai g èmmesa na m^joun th dom thc katanom c twn dedomènwn tou sunìlou ekpaÐdeushc kai ^ra mporoÔn na leitourg soun epituq¸c san Paragwgik^ Montèla [27], k^ti pou eÐqe arqik^ epeikonisteÐ sto sq ma 1 parap^nw.

SustolikoÐ (Contractive) Autìmatoi Kwdikopoihtèc

MÐa ^lloc mèjodoc gia kanonikopoÐhsh thc sun^rthshc kìstouc pou ja anafèroume gia lìgouc plhrìthtac eÐnai parìmoia me aut twn Arai¸n AK, dhl. h prìsjesh enìc

Sq ma 13: Sqhmatik apeikònish thc leitourgÐac twn Denoising Autìmatwn Kwdikopoiht¸n.
Phg : Anakataskeu apì Reconstruct corrupted data using Denoising Autoencoder (Medium Post), Garima Nishad, 2020, (https://medium.com/@analytics-vidhya/aeaff4b0958e)

ìrou sth sunˆrthsh aut . Sthn perÐptwsh twn Sustolik¸n AK, estiˆzoume sto pìso metabˆlletai h endiˆmesh anaparˆstash ìtan metablhjeÐ h eÐsodoc, me th logik ìti ìtan to dÐktuo tou kwdikopoiht (mèsw thc ekpaÐdeushc sunolikˆ tou SustolikoÔ AK) {mˆjei} ta basikˆ qarakthristikˆ kai th dom thc katanoom c twn dedomènwn, den ja prèpei h kwdikopoÐhsh thc eisìdou na metabˆlletai shmantikˆ me thn eÐsodo. Çtsi, o prìsjetoc ìroc thc sunˆrthshc kìstouc sÔmfwna me ton Rifai et al. [22], ja eÐnai:

$$(z; x) = \quad ! \quad \sum_i r_x z_i^2 \quad {}^2_F$$

Epomènwc h sunˆrthsh kìstouc gia th beltistopoÐhsh twn Sustolik¸n AK ja eÐnai:

$$L(x; f_{\tilde{}}; {}_{\tilde{}}g) = \text{ReconLoss}(x; g(f(x; {}_{\tilde{}}f); {}_{\tilde{}}g)) + \quad ! \quad \sum_i r_x z_i^2 \quad {}^2_F$$

kˆti pou enjarÔnei to montèlo na mˆjei mia sunˆrthsh pou den allˆzei polÔ gia gia mikrèc allagèc thc eisìdou (isodÔnamo me thn enjˆrrunsh tou montèlou na mˆjei th dom thc pijanotik c katanom c eisìdou) allˆ epÐshc na mporeÐ na anaktaskeuˆsei thn eÐsodo. Sth jewrÐa grammik¸n telest¸n, ènac grammikìc telest c lègetai {sustolikìc} eˆn ìtan efarmìzetai se eÐsodo monadiaÐac nìrmac, dÐnei èxodo me nìrma

mikrìterhc Ðshc thc monˆdac [70]. Çtsi, prosjèntontac ton ìro poin c thc Frobenius nìrmac tou Jacobian pÐnaka thc thc sunˆrthc tou kwdikopoiht , $z = f(x)|_{x=\bar{x}}$, timwroÔme to montèlo ìtan èqei megˆlec timèc ston Jacobian gia kˆje mÐa apì tic eisìdouc, dhl. ìtan h topik prosèggish katˆ Taylor thc $f(x)$ paÐrnei megˆlec timèc gia kˆpoia eÐsodo. Autì me th seirˆ tou pièzei touc epimèrouc topikoÔc grammikoÔc telestèc tou diktÔou na gÐnoun sustolikoÐ, exoÔ kai h onomasÐa aut c thc oikogèneiac AK.

MetablhtoÐ Autìmatoi Kwdikopoihtèc (MAK)

›swc h pio axios meiwth ideˆ stouc AK, h opoÐa èqei qrhsimopoihjeÐ ekten¸c kai gia Paragwgik MontelopoÐhsh eÐnai oi MetablhtoÐ Autìmatoi Kwdikopoihtèc (MAK) (Variational Autoencoders - VAEs), h dom kai oi efarmogèc twn opoÐwn analÔontai se aut n kai thn epìmenh upoenìthta. Sunoptikˆ, oi MAK eÐnai AK stouc opoÐouc h katanom pijanìthtac twn endiˆmeswn anaparastˆsewn qrhsimopoieÐtai sth kanoniko-poÐhsh(regularization) thc sunˆrthshc kìstouc me stìqo oi anaparˆstaseic autèc (ston lanjˆnonta pijanoq¸ro tou montèlou) na ekpaideujoÔn apokt¸ntac qr simec idiìthtec pou mporoÔn kat' epèktash na qrhsimopoihjoÔn gia thn paragwg nèwn dedomènwn.

Oi MAK an koun se mia eurÔterh oikogèneia pijanotik¸n montèlwn, aut twn Montèlwn Lanjˆnontwn Metablht¸n (Latent Variable Models). Sta montèla autˆ, mia sÔnjeth pijanotik katanom montelopoieÐtai mèsw enìc sunìlou lanjˆnouswn metablht¸n (latent variables), me tic opoÐec ta dedomèna eisìdou metasq matizontai ousiastikˆ se èna suneq dianusmatikì q¸ro mikrìterhc diˆstashc apì ton arqikì. Sugkekrimèna, dedomèna $\mathbf{X}$ ta opoÐa akoloujoÔn mia katanom $p_{data}(\mathbf{x})$, antistoiqÐzontai se lanjˆnousec metablhtèc $\mathbf{z}$ oi opoÐec me th seirˆ touc akoloujoÔn mia katanom , $p(\mathbf{z})$. H morf thc teleutaÐac orÐzetai apì prin kai gi' autì h $p(\mathbf{z})$ onomˆzetai prìterh (prior) katanom , se antidiastol me thn posterior katanom $p_{model}(\mathbf{z}|\mathbf{x})$ thn opoÐa proseggÐzoume stouc MAK ekpaideÔontac to TND tou kwdikopoiht . Tèloc, to dÐktuo tou apokwdikopoiht ekpaideÔetai gia na antistoiqÐzei tic lanjˆnousec metablhtèc pÐsw ston arqikì q¸ro, proseggÐzontac dhlad thn katanom $p_{model}(\mathbf{x}|\mathbf{z})$.

Stouc MAK, ta stoiqeÐa tou dianÔsmatoc thc endiˆmeshc anaparˆstashc, $\mathbf{z}$, apoteloÔn tic lanjˆnontec metablhtèc. H elpÐda eÐnai ìti efìson o kwdikopoiht c èqei katafèrei na mˆjei ènan metasqhmatismì se lanjˆnonta q¸ro ston opoÐo oi lanjˆnousec metablh-tèc antisoiqoÔn se sugkekrÐmena qarakthristikˆ parˆgontec diaforopoÐhshc (factors of variation) twn parathrÐsimwn metablht¸n (p.q. twn tim¸n twn eikonostoiqeÐwn gia

eÐsodo eikìna), tìte jewrhtikˆ {kinoÔmenoi} ston q¸ro autìn (dhl. kˆnontac tuqaÐec deigmatolhyeÐec apì th prior katanom p(z)) ja parˆgoume deÐgmata sthn èxodo tou apokwdikopoiht ta opoÐa an koun sthn katanom twn dedomènwn ekpaÐdeushc. Prokeimènou na epiteuqjeÐ autìc o skopìc, prèpei katˆ thn ekpaÐdeush twn dÔo diktÔwn na gÐnei h proanaferjeÐsa kanonikopoÐhsh wc proc to lanjˆnonta q¸ro. Stouc MAK gia na gÐnei autì arqikˆ antistoiqÐzoume kˆje eÐsodo se mÐa katanom sto lanjˆnonta q¸ro kai ìqi apl¸c se èna shmeÐo. Çtsi, to endiˆmeso diˆnusma z se èna MAK perièqei tic paramètrouc aut c thc katanom c gia thn ekˆstote eÐsodo, h opoÐa sun jwc orÐzetai na eÐnai Kanonik Isotropik (dhl. me diag¸nio pÐnaka summetablhtìthtac). Katìpin, deigmatolhptìume apì aut n thn pijanotik katanom , lambˆnoume èna diˆnusma z = z kai to zhtˆme apì ton apokwdikopoiht na anakataskeuˆsei thn arqik eÐsodo apì autì to deÐgma sto lanjˆnonta q¸ro, kˆti pou apeikonÐzetai grafikˆ sto sq ma 14 pou akolouje.

—pwc faÐnetai sto sq ma, h arqik eikìna pernˆei mèsa apì to dÐktuo tou kwdikopoiht h èxodoc tou opoÐou eÐnai oi parˆmetroi miac Poludiˆstathc Isotropik c Kanonik c katanom c. AkoloÔjwc arqikopoieÐtai mia tètoia katanom kai gÐnetai deigmatolhyÐa apì aut gia th paragwg thc eisìdou tou apokwdikopoiht . O apokwdikopoiht c kaleÐtai apì autì to deÐgma na anakataskeuˆsei me to elˆqisto sfˆlma thn arqik eikìna eisìdou. O kwdikopoiht c apì thn ˆllh kaleÐtai na mˆjei touc basikoÔc parˆgontec diaforopoÐhshc thc katanom c twn dedomènwn me skopì na h èxodìc tou na èqei ta basikˆ stoiqeÐa thc eisìdou. Paragwgikˆ montèla pou qrhsimopoioÔn touc (aploÔc) MAK teÐnoun na parˆgoun jolèc exìdouc lìgw thc je¸rhshc Kanonik c katanom c wc prìterhc.

H sunˆrthsh kìstouc epomènwc twn MAK ja perièqei ektìc apì ton ìro tou sfˆlmatoc anakataskeu c kai enan prìsjeto ìro, thn apìstash Kullback Leibler metaxÔ thc katanom c pou arqikopoieÐtai apì thn èxodo tou kwdikopoiht kai proseggÐzei thn posterior kai thc Ðdiac thc posterior kanonik c katanom c (pou ed¸ gia lìgouc aplìthtac dÐnetai na èqei mèsh tim 0 kai monadiaÐo pÐnaka summetablhtìthtac):

$$L(x; \tilde{f}; \tilde{g}) = \text{ReconLoss}(x; g(\text{Sample}[f(x; \tilde{f}); \tilde{g}])) + KL(N(\tilde{x}; \tilde{x}) || N(0; 1))$$

ìpou KL[] eÐnai h apìstash Kullback Leibler h opoÐa gia th perÐptwsh apìstashc metaxÔ

Sq ma 14: Apeikònish tou trìpou leitourgÐac kai sugkerkimèna tou emprìsjiou perˆsmatoc enìc MetablhtoÔ Autìmatou Kwdikopoiht .

Phg : Anakataskeu apì Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data , Teresa et al. , 2020 [111]

kanonik¸n katanom¸n èqei kleist morf [25]:

$$
\begin{aligned}
KL &= \int \left[\frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(x - \mu_1)^T \Sigma_1^{-1}(x - \mu_1) + \frac{1}{2}(x - \mu_2)^T \Sigma_2^{-1}(x - \mu_2) \right] p(x)dx \\
&= \frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2} \mathrm{tr}\left\{ E[(x - \mu_1)(x - \mu_1)^T] \Sigma_1^{-1} \right\} + \frac{1}{2} E[(x - \mu_2)^T \Sigma_2^{-1}(x - \mu_2)] \\
&= \frac{1}{2} \log \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2} \mathrm{tr}\{I_d\} + \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2) + \frac{1}{2} \mathrm{tr}\{\Sigma_2^{-1}\Sigma_1\} \\
&= \frac{1}{2} \left[\log \frac{|\Sigma_2|}{|\Sigma_1|} - d + \mathrm{tr}\{\Sigma_2^{-1}\Sigma_1\} + (\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1) \right] :
\end{aligned}
$$

ìpou d h diˆstash tou dianÔsmatoc endiˆmesh anaparˆstashc.

Autì pou ousiastikˆ epitugqˆnetai me th epibol miac sugkekrimènhc dom c gia th katanom twn endiˆmeswn anaparˆsteswn (enn. kanonik prìterh katanom) kai me th qr sh tou prìsjetou ìrou kanonikopoÐhshc aut c thc dom c eÐnai afenìc to ìti {kontinèc} shmasiologikˆ eÐsodoi {pièzontai} na antistoiqhjoÔn se kontinˆ shmeÐa ston lanjˆnonta q¸ro. Çtsi to montèlo pièzetai na mˆjei basikˆ qarakthristikˆ twn eisìdwn kai ìqi apl¸c na antigrˆfei thn eÐsodo sthn èxodo, en¸ tautìqrona mporoÔn na gÐnoun parambolèc metaxÔ shmeÐwn tou lanjˆnonta q¸rou kai autèc na apeikonistoÔn sthn èxodo wc omalèc parambolèc metaxÔ twn eisìdwn. Afetèrou, o lìgoc gia ton opoÐo qrhsimopoioÔntai kanonikèc katanomèc san priors eÐnai gia na eÐnai upologÐsimh

(tractable) h posterior, $p(z|x)$, kˆti pou alli¸c den ja Ðsque miac kai èqoume suneqeÐc metablhtèc gia thn $p(x)$ kai $p(z|x) = p(x|z)\,p(z) = p(x)$.

Paragwgik MontelopoÐhsh me MAK

Prin oloklhr¸soume thn parousiˆsh twn AK, jewroÔme skìpimo gia lìgouc plhrìthtac allˆ kai gia kalÔterh sÔndesh me ta Paragwgikˆ Montèla pou apoteloÔn to antikeÐmeno thc paroÔsac ergasÐac, na parajèsoume Ðswc to pio antiproswpeutikì paragwgikì montèlo to opoÐo qrhsimopoieÐ MAK. Prìkeitai gia to montèlo Vector-Quantized Variational Autoencoder (VQ-VAE)pou parousiˆsthke apì ton van den Oord et al. to 2017 [83] kai basÐzetai sth qr sh MAK twn opoÐwn ìmwc h proterh katanom den eÐnai stajer allˆ parametrik kai ekpaideÔsimh. EpÐshc, o kwdikopoiht c den bgˆzei suneqeÐc timèc allˆ diakritèc (dhl. o lanjˆnontac q¸roc den eÐnai suneq c allˆ apoteleÐtai apì memonomèna shmèia), mÐa èmpneush pou pro lje apì th jewrÐa thc dianusmatik c kbˆntishc (vector quantization) [7].

To montèlo VQ-VAEepekteÐnei ton tupikì AK prosjètwntac èna diakritì lexikì kwdik¸n codebook sto dÐktuo. To lexikì autì eÐna mia lÐsta dianusmˆtwn me antistoiqismènouc deÐktec kai qrhsimopoieÐtai gia kbˆntish tou dianÔsmatoc endiˆmeshc anaparˆstashc sthn èxodo tou diktÔou kwdikopoÐhshc. Çtsi, h èxodoc aut sugkrÐnetai me ìla ta dianÔsmata tou lexikoÔ kai san eÐsodo sto dÐktuo apokwdikopoÐshch epilègetai ekeÐno me th mikrìterh eukleÐdia apìstash:

$$z_q(x) = \underset{i}{\operatorname{argmin}} \|z_e(x) - e_i\|_2$$

ìpou $z_e(x)$ eÐnai h èxodoc tou kwdikopoiht gia eÐsodo x, e_i eÐnai to i-ostì diˆnusma tou lexikoÔ kai $z_q(x)$ eÐnai to kbantismèno diˆnusma to opoÐo pernˆei sthn eÐsodo tou apokwdikopoiht (bl. sq ma 15).

O kwdikopoiht c den bgˆzei èna mìno diˆnusma sthn èxodì tou allˆ èna sÔnolo tètoiwn dianusmˆtwn stoiqismèna p.q. gia efarmog se eÐsodo eikìnwn se èna plaÐsio 32×32, kajèna apì ta opoÐa kbantÐzetai anexˆrthta. Çtsi den tÐjetai jèma poikolìthtac (katˆrreushc thc katanom c - mode collapse - ìpwc alli¸c onomˆzetai) kaj¸c o apokwdikopoiht c èqei $|L|^B$ pijanèc eisìdouc, ìpou L to mègejoc tou lexikoÔ kai B to mègejoc tou plaisÐou.

Ta dianÔsmata tou lexikoÔ eÐnai kai ta Ðdia ekpaideÔsimec parˆmetroi tou montèlou.

Sq ma 15: Sqhmatik apeikìnish thc leitourgÐac tou montèlou VQ-VAE.
Phg : Anakataskeu apì Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data , Teresa et al. , 2020 [111]

EpÐshc h prìterh katanom aut¸n twn dianusmˆtwn eÐnai kai Ðdia ekpaideÔsimh. Gia thn ekpaÐdeush tou montèlou, h sunˆrthsh kìstouc perièqei ènan ìro me to sfˆlma anaktaskeus c kai dÔo akìmh ìrouc gia swst antistoÐqish twn exìdwn tou kwdikopoiht sto lexikì kai thn katˆ to dunatì kalÔterh suneisforˆ tou kwdikopoiht sthn ananèwsh tou lexikoÔ. Akìmh, gia thn ekpaÐdeush thc prìterhc katanom c oi dhmiourgoÐ dokÐmasan AutopalindronoÔmena montèla kai sugkekrimèna to PixelCNN me arketˆ entupwsiakˆ apotelèsmata.

(a) Eikìnec apì to ImageNET 128Ö128 (b) Anakataskeu apì VQ-VAEmeL=512 kai B=32

Sq ma 16: Efarmog montèlou VQ-VAE se eikìnec apì to sÔnolo dedomènwn ImageNET. O realismìc kai h poikilÐa twn paragìmenwn eikìnwn eÐnai pragmatikˆ axioshmeÐwta.
Phg : Neural Discrete Representation Learning , van den Oord et al., 2017 [83]

Tèloc, h parousiˆsh den ja tan pl rhc eˆn den anafèrame ìti mia exèlixh tou VQ-VAEh opìia qrhsimopoieÐ ENN gia ekpaÐdeush thc prior (enn. afoÔ pr¸ta èqei qrhsimopoihjeÐ

mia stajer gia thn ekpaÐdeush tou montèlou), eÐnai to montèlo DALL-E pou parousi-
ˆsthke to 2021 apì Ramesh et al. thc OpenAI[121] kai apoteleÐ to pio exeligmèno montèlo metatrop c keimènou se eikìna èwc s mera (IoÔnioc 2021). Sto sq ma 17 pou akolouje faÐnontai paragwgèc tou montèlou pou epideiknÔoun thn apìluth uperoq tou.

Sq ma 17: Paragwgèc tou montèlou DALL-E thc OpenAI gia eÐsodo th frˆsh an armchair in the shape of an avocado . Oi eikìnec den qreiˆzontai peraitèrw sqoliasmì.
Phg : DALL-E: Creating Images from Text (OpenAI Blog), Ramesh et al., 2021[120]

2.4 Paragwgik MontelopoÐhsh me Generative Adversarial Networks

Metˆ apì thn eisagwg se ˆllec morfèc montèlwn kai teqnik¸n gia Paragwgik Montelopo-
Ðhsh, epanerqìmaste sto tèloc tou 1ou kefalaÐou se mia sÔntomh eisagwg thc dom c kai leitourgÐac twn Paragwgik¸n Antiparajetik¸n DiktÔwn Generative Adversarial Networks (GANs) - ja ta onomˆzoume GANs gia suntomÐa - sta opoÐa epikentrwnìmste èwc to tèloc thc paroÔsac ergasÐac. Ta GANs ìpwc proanafèrjhke, parousiˆsthkan apì ton Ian Goodfellow et al. to 2014 [29] mia perÐodo pou h Paragwgik MontelopoÐhsh eikìnwn (me thn opoÐa asqol jhkan oi dhmiourgoÐ touc) ginìntan kurÐwc me Autopalindro-
nouÔmena montèla montèla AK. Aform thc anaz thshc nèwn teqnik¸n apotèlese to gegonìc ìti h paragwgik dunatìthta exartoÔntan èntona sta AutopalindronouÔmena montèla apì th seirˆ paragwg c twn eikonostoiqeÐwn, en¸ stouc AK apì th upìjesh prìterhc katanom c kai th dom aut c. Phg èmpneushc, epÐshc, apotèlesan oi melètec kai prospˆjeiec gia antiparajetik ekpaÐdeush (adversarial training) TND, kaj¸c eÐqe brejeÐ pwc h uperbolik qr sh grammik¸n sunˆrthsewn odhgeÐ èna montèlo me TND na èqei euaisjhsÐa se metabolèc thc eisìdou (en¸ se antÐjesh ja tan epijumhtì na upˆrqei sqetikˆ stajer sumperiforˆ tou montèlou gÔrw apì ta deÐgmata tou sunìlou ekpaÐdeushc) [30].

Basikì stoiqeÐo thc dìmhc twn GANs, to opoÐo ta diaforopoieÐ apì ta prohgoÔmena

BPM, eÐnai h Ôparxh kai tautìqronh ekpaÐdeush dÔo diktÔwn: tou Generator pou dhmiourgeÐ deÐgmata katˆ to dunatì kontinìtera se autˆ tou sunìlou ekpaÐdeushc kai tou Discriminator pou ekpaideÔetai na xeqwrÐzei poia deÐgmata proèrqontai apì to sÔnolo ekpaÐdeushc (dhl. eÐnai ta pragmatikˆ - real) kai poia apì ton Generator (dhl. eÐnai ta teqnhtˆ yeÔtika - fake). Sugkekrimèna, se kˆje b ma ekpaÐdeushc (dhl. sto eswterikì tou brìgqou ekpaÐdeushc - sq ma 18) o Discriminator, pou an kei sth kathgorÐa twn Diakritik¸n Montèlwn (bl. enìthta 2.1), dèqetai deÐgmata apì to sÔnolo dedemènwn ekpaÐdeushc kai deÐgmata pou èqoun paraqjeÐ apì ton Generator kai ekpaideÔetai na bgˆzei pijanìthta kontˆ sto 1 gia ta pr¸ta kai kontˆ sta 0 gia ta deÔtera, en¸ o Generator pou an kei sth kathgorÐa twn Paragwgik¸n Montèlwn ekpaideÔetai ¸ste apì eÐsodo jìrubo na bgˆzei eikìnec sthn èxodì arketˆ realistikèc ¸ste na mporèsoun na {xegelˆsoun} ton Discriminator.

EmbajÔnontac lÐgo sthn anˆlush tou trìpou ekpaÐdeushc twn GANs, mporoÔme na poÔme ìti tìso o Generator ìso kai o Discriminator antiproswpeÔontai apì (suneq¸c) diaforÐsimec sunart seic me ekpaideÔsimec paramètrouc, ìpwc eÐnai ta TND, kˆje mÐa me th dik thc sunˆrthsh kìstouc. Ta dÔo dÐktua ekpaideÔontai mèsw back-propagation qrhsimopoi¸ntac amfìtera th sunˆrthsh kìstouc tou Discriminator, allˆ me diaforetikì stìqo. O Discriminator prospajeÐ na mei¸sei th sunˆrthsh kìstouc tìso gia ta pragmatikˆ ìso kai gia ta teqnhtˆ deÐgmata, en¸ o Generator prospajeÐ na aux sei th sunˆrthsh kìstouc tou Discriminator gia ta teqnhtˆ deÐgmata pou parˆgei. Axioshmei¸to eÐnai, epiprìsjeta, ìti to sÔnolo dedomènwn ekpaÐdeushc kai mìno autì kajorÐzei to eÐdoc twn deigmˆtwn pou o Generator majaÐnei na parˆgei. Çtsi, eˆn gia parˆdeigma epijumoÔme èna paragwgikì montèlo tÔpou GAN na parˆgei realistikèc eikìnec apì topÐa (p.q. gia efarmogèc anabˆjmishc realismoÔ twn frames enìc paiqnidioÔ - bl. NVIDIA DLSS® [109] kai AMD FidelityFX® [118]), ja prèpei to sÔnolo dedomènwn ekpaÐdeushc na apoteleÐtai apì zeÔgh qamhl c-uyhl c anˆlushc eikìnwn tètoiwn topÐwn. Autì eloqeÔei èna shmantikì kÐnduno: mh-leptomer¸c kataskeuasmèna sÔnola dedomènwn ekpaÐdeushc mporeÐ na odhg soun BPM tÔpou GAN na parousiˆzoun èntonh merolhyÐa, kˆti pou telikˆ mei¸nei thn axiopistÐa kai qrhsimìthtˆ touc [80], [101].

Sq ma 18: Grafik anaparˆstash tou eswterikoÔ brìgqou ekpaÐdeushc enìc GAN. O Discrimina-tor lambˆnei eÐte pragmatikèc eikìnec apì to sÔnolo ekpaÐdeushc eikìnec pou èqoun paraqjeÐ apì ton Generator kai bgˆzei èna skor {realistikìthtac} gia kˆje mÐa. Ta skìr qrhsimopoioÔntai akoloujiakˆ apì kˆje dÐqtuo gia ananèwsh twn ekˆstote paramètrwn.
Phg : Anakataskeu apì GANs in action: deep learning with generative adversarial networks , Langr et al., 2019[104]

Prin thn olokl rwsh tou parìntoc kefalaÐou, jewroÔme skìpimo na anafèroume epigrammatikˆ merikèc peript¸seic ìpou ta GANs efarmìsthkan jewroÔme pwc ja efarmostoÔn me epituqÐa. Arqikˆ, h montelopoÐhsh qarakthristik¸n anjr¸pinwn pros¸pwn ektoxeÔjhke me th qr sh GANs kˆti pou apeikonÐzetai sto sq ma 2 sthn arq tou kefalaÐou. Gia th montelopoÐhsh aut wc epÐ to pleÐsto qrhshmiopoeÐtai to sÔnolo dedomènwnFlickr-Faces-HQ Dataset (FFHQ)to opoÐo parousiˆsthke mazÐ me to state-of-the-art montèlo StyleGAN[99], [101]. 'Allo entupwsiakì parˆdeigma eÐnai h metatrop enìc skÐtsou se mÐa fwtorealistik apeikìnish kai mˆlista se pragmatikì qrìno, me to montèlo GauGANpou analÔetai sthn enìthta 4.2.2 parakˆtw. Arketèc prospˆjeiec, akìmh, gÐnontai kai apì thn etaireÐa Adobe, me to tm ta èreunac aut c na sqediˆzei thn epìmenh geniˆ tou Photoshop® me enswmˆtwsh GANs gia diìrjwsh [93] sÔnjesh eikìnwn [95], en¸, sta plaÐsia thc montelopoÐhshc eikìnwn mìdac, apì to Ðdio tm ma èqei parousiasteÐ Ðswc to pio apodotikì montèlo gia eikonikì dokimast rio (virtual try-on), to SieveNet[112], parˆdeigma efarmog c tou opoÐou faÐnetai sto sq ma 19 pou akoloujeÐ.

(a) Arqik eikìna (b) RoÔqo gia dokim (g) Çxodoc montèlou dokim c

Sq ma 19: Paragwg tou montèlou SieveNet thc Adobe gia eÐsodo thn arqik eikìna tou {doki-mast } (aristerˆ), kaj¸c kai tou roÔqou proc dokim (kèntro) kai èxodo thn eikìna sta me to dokimast na forˆei to roÔqo-stìqo (dexiˆ). To SieveNetjewreÐtai to pio kainotìmo montèlo gia virtual try-on, en¸ qrhsimopoieÐ kai GANs.

Phg : SieveNet: A Uni ed Framework for Robust Image-Based Virtual Try-On , Jandial et al., 2020 [112]

Kefˆlaio 3

Ekpaideush twn GANs

Sto kefˆlaio autì ja embajÔnoume thn anˆlus mac sqetikˆ me ton trìpo dìmhshc kai ekpaÐdeushc BPM tÔpou GAN, ja anaferjoÔme sta diˆfora sthsÐmata (setups) pou èqoun parousiasteÐ gia thn apotelesmatikìterh ekpaÐdeush tètoiwn montèlwn, en¸ sto tèloc ja estiˆsoume sthn axiolìghsh twn paragwg¸n apì GANs - ìla sta plaÐsia Paragwgik c MontelopoÐhshc eikìnwn. —pwc anafèrjhke kai sthn anˆlush pou prohg jhke, ta GANs apoteloÔntai tupikˆ apì dÔo (diakritˆ) neurwnikˆ dÐktua, kajèna me th dik tou sunˆrthsh kìstouc gia beltistopoÐhsh twn dik¸n tou ekpaide-Ôsimwn paramètrwn (bar¸n, pol¸sewn, treqìntwn statistik¸n klp). Autì, ìpwc ja analujeÐ stic parakˆtw enìthtec, duskoleÔei shmantikˆ thn eustaj kai epituqhmènh ekpaÐdeush enìc GAN, wstìso apoteleÐ Ðswc to kuriìtero qarakthristikì gia thn epituqÐa kai dhmofilÐa touc.

Stic akìloujec enìthtec ja analÔsoume ton trìpo ekpaÐdeushc twn GANs, xekin¸ntac apì mÐa sÔgkrish autoÔ me ton trìpo ekpaÐdeushc twn ˆmeswn progìnwn touc, twn AK. Sth sunèqeia ja parousiˆsoume tic diˆforec morfèc sqhmatismoÔ thc sunˆrthshc kìstouc kai ekpaÐdeushc twn GANs, en¸ sto tèloc tou kefalaÐou ja anaferjoÔme se teqnikèc axiolìghshc twn paragìmenwn eikìnwn apì autˆ. PrwtoÔ suneqÐsoume, wstìso, jewroÔme skìpimo gia lìgouc plhrìthtac na perigrˆyoume kai ed¸ ta basikˆ domikˆ stoiqeÐa enìc GAN: to dÐktuo diˆkrishc Discriminator kai to dÐktuo paragwg c Generator. Shmei¸noume sto shmeÐo autì pwc gia to upìloipo thc paroÔsac ergasÐac ja anaferìmaste sta dÐktua autˆ me tou agglikoÔc ìrouc gia lìgouc aplìthtac kai eukolìterhc anˆgnwshc.

DÐktuo Diˆkrishc, Discriminator

To pr¸to domikì mèroc enìc GAN eÐnai to dÐktuo diˆkrishc Discriminator dÐktuo ìpwc onomˆzetai. Autì eÐnai èna tupikì dÐktuo taxinomht ìpwc autˆ pou parousiˆsthkan sta diakritikˆ montèla (bl. enìthta 2.1). Sta plaÐsia thc Paragwgik c MontelopoÐhshc eikìnwn oi Discriminators apoteloÔntai apì SuneliktikˆNeurwnikˆ DÐktua, ìpwc p.q. o Discriminator tou montèlou DCGAN pou faÐnetai dexiˆ kai opoÐoc apoteleÐtai apì suneliktikèc str¸seic h mia metˆ thn ˆllh kai mia pl rwc sundedemènh str¸sh prin thn èxodo gia exagwg thc pijanìthtac realismoÔ thc eisìdou, $p(\text{real}|x)$. Prìkeitai, epomènwc, gia dÐktua duadik¸n taxinomht¸n eikìnac . Genikˆ, epeid to èrgo twn Discriminators sta GANs eÐnai arketˆ pio aplì se sqèsh me autì twn taxinomht¸n eikìnwn (gia parˆdeigma tou ImageNET) pou anafèrjhkan sto prohgoÔmeno kefˆlaio, oi pr¸toi teÐnoun na eÐnai pio aplˆ dÐktua me shmantikˆ ligìterec ekpaideÔsimec paramètrouc apì ta teleutaÐa.

Sq ma 20: Discriminator tou DCGAN(bl. 4.1.1)
Phg : Semantic Image Inpainting with Perceptual and Contextual Losses , Yeh et al., 2016 [65]

DÐktuo Paragwg c, Generator

To deÔtero domikì mèroc enìc GAN eÐnai to dÐktuo paragwg c, Generator dÐktuo ìpwc onomˆzetai. Sthn perÐptwsh aut , to dÐktuo ekteleÐ èrgo antÐstrofo apì autì tou Discriminator , kaj¸c o stìqoc ed¸ eÐnai h èxodoc na eÐnai eikìna kai mˆlista mÐa pou parousiˆzei realismì kai poikilomorfÐa metaxÔ twn upìloipwn eikìnwn pou parˆgontai. Sthn apl tou morf , wc eÐsodo ston Generator dÐnetai èna diˆnusma leukoÔ jorÔbou (sun jwc gkaousianoÔ), to opoÐo pernˆei mèsa apì mia akoloujÐa anestrammènwn suneliktik¸n str¸sewn (transposed convolutions) . Parˆdeigma tètoiou Generator faÐnetai epˆnw dexiˆ apì to montèlo DCGAN¸ to opoÐo tan kai

Sq ma 21:Generator tou DCGAN(bl. 4.1.1)
Phg : Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks , Radford et al. , 2016 [44]

to pr¸to pou qrhsimopoioÔse tètoiec str¸seic gia ton Generator - dom pou plèon jewreÐtai epikratoÔsa. Stìqoc, epomènwc thc ekpaÐdeushc tou Generator eÐnai na mˆjei èmmesa thn katanom pijanìthtac twn eikìnwn (enn. thn poludiˆstath katanom me diastˆseic ta eikonostoiqeÐa thc eikìnac exìdou) tou sunìlou dedomènwn, $p(x)$, th desmeumènh èkdosh aut c sth perÐptwsh pou sthn eÐsodo tou Generator upˆrqei kai kˆpoia sunj kh, $p(x|y)$ (bl. 3.2).

3.1 Sunart seic Kìstouc kai PaÐgnia

SÔgkrish GANs me Autìmatouc Kwdikopoihtèc

H antiparabol twn GANs me touc AK èrqetai fusikˆ. Ki autì diìti, ìpwc faÐnetai kai sto sq ma 22 parakˆtw, o Generator kˆnei parìmoia douleiˆ me aut n tou diktÔou apokwdikopoÐhshc stouc AK: lambˆnontac sthn eÐsodì tou èna diˆnusma, prospajeÐ na parˆxei dedomèna ta opoÐa den xeqwrÐzoun apì autˆ tou sunìlou ekpaÐdeushc. EpÐshc, o Discriminator lambˆnei eikìna sthn eÐsodì tou, ìpwc antÐstoiqa gÐnetai kai sto dÐktuo kwdikopoÐhshc enìc AK.

Wstìso, parousiˆzoun shmantikèc diaforèc, me th shmantikìterh na entopÐzetai ston trìpo upologismoÔ thc sunˆrthshc kìstouc gia beltistopoÐhsh twn paramètrwn tou ekˆstote montèlou. Çtsi, en¸ stouc AK aut sun jwc {metrˆei} to sfˆlma anakataskeu c miac arqik c eikìnac apì thn endiˆmesh anaparˆstash z, sta GANs qrhsimopoieÐtai èna prìsjeto neurwnikì dÐktuo pou ekpaideÔetai na diaqwrÐzei ta alhjinˆ apì ta teqnhtˆ dedomèna, sqhmatÐzontac ètsi kai th sunˆrthsh kìstouc tou Generator pou metrˆei to bajmì realismoÔ twn deigmˆtwn pou parˆgei. Çtsi, ìpwc faÐnetai kai sto parakˆtw sq ma, stouc MAK, èrgo tou apokwdikopoiht (decoder) eÐnai na anakataskeuˆsei thn eÐsodo tou kwdikopoiht (encoder) apì deÐgma mia ekpaideÔsimhc katanom c, kˆti ˆmesa metr simo apì to antÐstoiqo sfˆlma. Sta GANs, o Generator prèpei na mˆjei na kataskeuˆzei eikìnec mh diakrÐsimec apì autèc tou sunìlou ekpaÐdeushc kˆnontac qr sh ìmwc enìc ˆllou neurwnikoÔ diktÔou, tou Discriminator, gia mètrhsh thc apotelesmatikìthtˆc tou, dhlad majaÐnontac èmmesa pìso kalˆ ta phgaÐnei. Tèloc, sta GANsta dÔo dÐktua antimˆqontai, to èna gia na mei¸sei ta sfˆlmata taxinìmhshc wc proc to realismì (Discriminator) kai to ˆllo gia na ta aux sei stic eikìnec pou parˆgei (Generator).

Mia akìmh diaforopoÐhsh twn dÔo teqnik¸n, ègkeitai sto gegonìc ìti oi AK eÐnai anti-strèyima dÐktua, me thn ènnoia ìti gia dedomènh eikìna exìdou tou apokwdikopoiht , o kwdikopoiht c ja mac d¸sei thn endiˆmesh anaparˆstash apì thn opoÐa proèkuye (akìma ki an prìkeitai gia paramètrouc katanom c, ìpwc stouc MAK). Sta GANswstìso kˆti tètoio den eÐnai ˆmesa efiktì, kaj¸c sun jwc gÐnetai mia tuqaÐa deigmatolhyÐa (noise-to-image) kˆpoioc mh antistrèyimoc metasqhmatismìc (image-to-image) sthn eÐsodo. Tèloc, mia ligìtero kaÐria diaforˆ eÐnai ìti o Discriminator se èna GAN pa-Ðrnei eikìna kai bgˆzei (toulˆqiston sth morf pou protˆjhkan arqikˆ) ènan arijmì pou sqetÐzetai me to bajmì realismoÔ thc eisìdou, en¸ to dÐktuo tou Kwdikopoiht (decoder) enìc AK gia eÐsodo eikìna epistrèfei èna diˆnusma ston lanjˆnonta q¸ro tou montèlou.

Sq ma 22: SÔgkrish thc dom c kai tou trìpou ekpaÐdeushc enìc GAN me èna montèlo MAK (VAE). Phg : Anakataskeu apì Flow-based Deep Generative Models , Lilian Weng, 2018[98]

Sqhmatismìc thc sunˆrthshc kìstouc enìc GAN

H sunˆrthsh kìstouc tou kajenìc apì ta epimèrouc dÐktua enìc GAN, se antÐjesh me ˆlla BPN kai Diakritikˆ Montèla me TND, den èqei wc paramètrouc mìno autèc tou

autoÔ diktÔou, allˆ kai tou ˆllou. Sugkekrimèna, akolouj¸ntac thn tupik onomatodosÐa, èstw θ_G oi ekpaideÔsimec parˆmetroi tou Generator diktÔou kai θ_D tou Discriminator. AntÐstoiqa, h sunˆrthsh kìstouc ja eÐnai J_G gia ton Generator kai J_D gia ton Discriminator, me kˆje dÐktuo na prospajeÐ na beltiˇsei th dik tou (pijanìn kai me diaforetikoÔc algorÐjmouc beltistopoÐhshc). Epomènwc, perifrastikˆ oi sunart seic kìstouc twn diktÔwn enìc GAN ja eÐnai [50]:

$J_D(\theta_G; \theta_D)$ = elaqistopoÐhsh tou sfˆlmatoc taxinìmhshc se pragmatikˆ kai teqnhtˆ dedomèna

$J_G(\theta_D; \theta_G)$ = megistopoÐhsh tou sfˆlmatoc taxinìmhshc tou Discriminator sta paragìmena dedomèna

Parìti, ìmwc, h kˆje sunˆrthsh kìstouc emplèkei kai paramètrouc tou ˆllou diktÔou, qrhsimopoieÐtai gia na ananeˇsei mìno autèc tou dikoÔ thc. Çtsi, o algìrijmoc beltistopoÐhshc tou Discriminator, gia parˆdeigma, qrhsimopoieÐ thn $J_D(\theta_D; \theta_G)$ kai thc merikèc paragˇgouc aut c wc proc tic paramètrouc tou Discriminator gia thn ekpaÐdeush twn paramètrwn tou Discriminator, θ_D, qwrÐc na mporeÐ na ephreˆsei tic paramètrouc tou Generator (tic merikèc paragˇgouc autˇn). AntÐstoiqa gia to dÐktuo tou Generator kai ton algìrijmo beltistopoÐhshc autoÔ. Dhlad se kˆje èna apì ta dÔo stˆdia tou eswterikoÔ brìgqou ekpaÐdeushc enìc GAN pou apeikonÐzetai sto sq ma 18, to èna apì ta dÔo dÐktua paramènei {pagwmèno} (frozen).

—pwc gÐnetai emfanèc apì ta parapˆnw ta dÔo dÐktua {antimˆqontai} to èna to ˆllo, kˆti pou an kei sthn eurÔterh oikogèneia teqnikˇn antiparajetik c ekpaÐdeushc (adversarial training). Sto tèloc, ta teqnhtˆ deÐgmata pou parˆgontai apì ton Generator faÐnontai tìso alhjinˆ ston Discriminator pou to kalÔtero pou èqei na kˆnei eÐnai na mantèyei sth tÔqh se poia omˆda deigmˆtwn an koun. Tìte, lème ìti h ekpaÐdeush èqei oloklhrwjeÐ epituqˇc kai plèon mporoÔme na qrhsimopoi soume ton ekpaideumènoGenerator gia na parˆgoume realistikˆ dedomèna parìmoia me autˆ tou sunìlou ekpaÐdeushc. Wstìso, gia na ftˆsoume sto shmeÐo autì, prèpei na rujmÐsoume ton trìpo upologismoÔ thc sunˆrthshc kìstouc gia kˆje dÐktuo kai pwc autèc upologÐzontai apì ta apotelèsmata thc taxinìmhshc sthn èxodo tou Discriminator, kˆti pou analÔetai stic paragrˆfouc pou akoloujoÔn.

PaÐgnio MhdenikoÔ AjroÐsmatoc - Binary Cross-Entropy

Epeid o Generator kai o Discriminator mporoÔn na beltiˌsoun mìno tic dikèc tou paramètrouc ìqi o ènac tou ˆllou, h ekpaÐdeush enìc GAN mporeÐ kalÔtera na perigrafeÐ wc èna paÐgnio. Oi paÐktec autoÔ tou paignÐou eÐnai ta dÔo neurwnikˆ dÐktua kai ìso to èna gÐnetai kalÔtero, tìso to ˆllo qeirotereÔei kai prospajeÐ ek' nèou na beltiwjeÐ ˌste na prosperˆsei to prˌto. Sta plaÐsia thc JewrÐac PaignÐwn, èna tètoio st simo (setup) eÐnai gnwstì wc paÐgnio mhdenikoÔ ajroÐsmatoc dÔo paiktˌn (two-player zero-sum game), ìpou ta kèrdh tou enìc paÐkth isoÔntai me tic apˌleiec tou ˆllou , antÐstoiqa, h beltÐwsh enìc paÐkth katˆ èna pososthì isodunameÐ me qeirotèreush tou ˆllou katˆ to Ðdio pososthì.

Se ìla ta paÐgnia mhdenikoÔ ajroÐsmatoc upˆrqei èna shmeÐo isorropÐac (mia minimax[1] lÔsh ìpwc lègetai), gnwstì wc Nash equilibrium, enìc shmeÐou sto opoÐo kanènac paÐkthc den mporeÐ na beltiˌsei thn katˆstas tic apodìseic tou allˆzontac tic enèrgeièc tou. Autì sumbaÐnei sta GANs, ìtan ìpwc anafèrjhke kai prohgoÔmena, o Generator parˆgei realistikèc eikìnec mh-diakrÐsimec apì autèc tou sunìlou ekpaÐdeushc kai o Discriminator sthn kalÔterh mporeÐ na mantèyei tuqaÐa eˆn èna parˆdeigma dedomènwn sthn eÐsodì tou proèrqetai apì to pragmatikì sÔnolo dedomènwn èqei paraqjeÐ apì ton Generator [50]. Tìte lègetai ìti to GANèqei sugklÐnei kai h ekpaÐdeus tou oloklhrˌnetai, kˆti pou sthn prˆxh eÐnai polÔ dÔskolo na epiteuqjeÐ, qwrÐc ìmwc autì na empodÐzei thn epituq efarmog tètoiwn montèlwn sthn prˆxh [104].

Apì mÐa pio majhmatik skopiˆ, to Nash equilibrium sta GANs epitugqˆnetai ìtan h sunˆrthsh kìstouc tou Generator, $J_G\ \tilde{}_G; \tilde{}_D$ èqei elaqistopoihjeÐ wc proc tic dikèc tou ekpaideÔsimec paramètrouc, $\tilde{}_G$ kai tautìqrona h sunˆrthsh kìstouc tou Discriminator, $J_D\ \tilde{}_D; \tilde{}_G$ èqei elaqistopoihjeÐ wc proc tic dikèc tou ekpaideÔsimec paramètrouc, $\tilde{}_D$, kˆti pou apeikonÐzetai grafikˆ sto sq ma 23 parakˆtw: o paÐkthc 1 (Discriminator) prospajeÐ na meiˌsei th sunˆrthsh kìstouc, V, beltistopoiˌntac tic paramètrouc tou, $_1$, enˌ o paÐkthc 2 prospajeÐ na meiˌsei th dik tou sunˆrthsh kìstouc, V (dhl. na megistopoi sei th V) beltistopoiˌntac tic dikèc tou paramètrouc, $_2$. H sunolik sunˆrthsh kìstouc faÐnetai sto plegmatoeidèc sq ma dexiˆ tou qˌrou twn sunolikˌn paramètrwn twn dÔo diktÔwn. EpÐshc se autì to sq ma, faÐnetai (maÔrh gramm me

[1]Sth JewrÐa PaignÐwn, minimax tim enìc paÐkth eÐnai h mikrìterh tim pou oi ˆlloi paÐktec mporoÔn na anagkˆsoun ton paÐkth na lˆbei, qwrÐc na gnwrÐzoun tic enèrgeièc tou, isodÔnama, eÐnai h megalÔterh exasfalismènh axÐa pou mporeÐ na pˆrei o paÐkthc ìtan xèrei tic enèrgeiec twn ˆllwn paiktˌn.

teleÐec) to monopˆti sÔgklishc sto Nash equilibrium perÐpou sto kèntro autoÔ, sto diˆselo (saddle-point)

Sthn arqik èkdosh twn GANs [29], wc sunˆrthsh upologismoÔ tou kìstouc gia kˆje dÐktuo eÐqe protajeÐ h Duadik DiastauroÔmenh EntropÐa (Binary Cross-Entropy - BCE), h opoÐa kai efarmìzetai se ergasÐec taxinìmhshc ìpou upˆrqoun dÔo kathgorÐec, ìpwc h ergasÐa pou kˆnei o Discriminator. —pwc analÔetai sth sunèqeia, autìc den eÐnai o mìnoc trìpoc upologismoÔ tou kìstouc tou kˆje diktÔou kai mˆlista o Ðswc ligìtero qrhsimopoioÔmenoc lìgw tou koresmoÔ (saturation) kai kat' epèktash h exafˆnish twn parag¸gwn (vanishing gradients) pou prokaleÐ h qr sh logarÐjmwn.

Sq ma 23: Sqhmatik apeikìnish thc ekpaÐdeushc enìc GAN wc èna paÐgnio mhdenikoÔ ajroÐsmatoc.
Phg : Anakataskeu apì Goodfellow, 2019, https://www.iangoodfellow.com/slides/2019-05-07.pdf

H sunˆrthsh kìstouc Binary Cross-Entropy

H sunˆrthsh kìstouc Binary Cross-Entropygia èna sÔnolo m deigmˆtwn anˆ omˆda (batch) èqei wc ex c:

$$J_m(\tilde{}) = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h(x^{(i)};\tilde{}) + (1 - y^{(i)}) \log (1 - h(x^{(i)};\tilde{})) \quad (3.1)$$

ìpou to arqikì ˆjroisma kai h diaÐresh me ton arijmì twn deigmˆtwn proseggÐzei ton telest mèshc tim c, $x^{(i)}$ eÐnai to i-ostì deÐgma, $y^{(i)}$ eÐnai h etikèta tou i-ostoÔ deÐgmatoc kai $\tilde{}$ to diˆnusma twn ekpaideÔsimwn paramètrwn tou montèlou. Katˆ thn ekpaÐdeush tou Discriminator enìc GAN oi etikètec ja eÐnai 1 gia ta pragmatikˆ deÐgmata kai 0 gia ta teqnhtˆ, en¸ gia thn ekpaÐdeush tou Generator isqÔei to antÐstrofo, dhlad mazÐ me ta

teqnhtˆ deÐgmata ja dÐnetai h etikèta 1 prokeimènou na upologÐsei to katˆ pìso mporeÐ na {xegelˆsei} ton Discriminator.

Estiˆzontac sto sqhmatismì thc sunˆrthshc kìstouc kai stic timèc pou aut lambˆnei gia tic etikètec 0=1 pou dÐnontai katˆ thn ekpaÐdeush enìc GAN, blèpoume ìti ìtan h etikèta eÐnai 1 dra mìno o pr¸toc ìroc tou ajroÐsmatoc, opoÐoc paÐrnei timèc apì to 0 (ìtan h $h\ x^{(i)};\sim$ bgˆzei timèc kontˆ sto 1) èwc to 1 (ìtan h $h\ x^{(i)};\sim$ bgˆzei timèc kontˆ sto 0). AntÐstoiqa, gia etikèta 0, dra mìno o deÔteroc ìroc tou ajroÐsmatoc, opoÐoc paÐrnei timèc apì to 0 (ìtan h $h\ x^{(i)};\sim$ bgˆzei timèc kontˆ sto 0) èwc to 1 (ìtan h $h\ x^{(i)};\sim$ bgˆzei timèc kontˆ sto 1). Lambˆnontac, tèloc, upìyh kai to arnhtikì prìshmo sthn arq thc 3.1, blèpoume ìti h parapˆnw prosèggish thc Binary Cross-Entropygia èna batch paÐrnei timèc apì 0 èwc +1 ìtan h sunˆrthsh taxinìmhshc, h(x) me paramètrouc tic paÐrnei timèc apì to 0 èwc to 1. Epilogikˆ, h sunˆrthsh kìstouc Binary Cross-Entropy, èqei dÔo mèrh (èna gia kˆje tˆxh) kai lambˆnei timèc kontˆ sto 0 gia swst taxinìmhsh (diag¸nioc confusion matrix) en¸ proseggÐzei to jetikì ˆpeiro gia lˆjoc (antidiag¸nioc confusion matrix) - sumperiforˆ pou apeikonÐzetai grafikˆ sto sq ma 24 parakˆtw.

(a) Gia etikèta $y^{(i)} = 1$ (b) Gia etikèta $y^{(i)} = 0$

Sq ma 24: Grafik anaparˆstash thc sunˆrthshc kìstouc Binary Cross-Entropy gia etikètec 1 (pragmatik¸n eikìnwn) aristerˆ kai 0 (teqnht¸n) dexiˆ.
Phg : Anakataskeu apì Generative Adversarial Networks Specialization, Zhou et al., DeepLearning.AI, 2021[122]

Sqhmatismìc Sunart sewn Kìstouc se èna GAN

Me bˆsh thn anˆlush pou prohg jhke, eÐmaste plèon se jèsh na grˆyoume se kleist morf th sunˆrthsh kìstouc pou kaleÐtai na beltistopoi sei to kˆje dÐktuo. Çtsi, gia ton Discriminator, h sunˆrthsh kìstouc eˆn qrhsimopoihjeÐ h Binary Cross-Entropykai me dedomèno ìti katˆ thn ekpaÐdeus enìc GANta pragmatikˆ dedomèna eÐnai katˆ sÔmbash

antistoiqismèna me thn etikèta 1 kai ta teqnhtˆ me thn 0, ja eÐnai:

$$J_D(\tilde{\theta}_D, \tilde{\theta}_G) = \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log h(x^{(i)};\tilde{\theta}) + (1-y^{(i)})\log(1-h(x^{(i)};\tilde{\theta}))\right] \quad (3.2)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[\log D(x^{(i)};\tilde{\theta}_D) + \log(1-D(G(z^{(i)};\tilde{\theta}_G);\tilde{\theta}_D))\right] \quad (3.3)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\log D(x^{(i)};\tilde{\theta}_D) - \frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^{(i)};\tilde{\theta}_G);\tilde{\theta}_D)) \quad (3.4)$$

$$E_{x\sim p_{data}}\log[D(x)] - E_{z\sim p_{prior}}\log[1-D(G(z))] \quad (3.5)$$

ìpou $D(x)$ eÐnai h èxodoc Discriminator (dhl. h pijanìthta realismoÔ thc eisìdou x), $G(z)$ eÐnai h èxodoc tou Generator diktÔou gia eÐsodo tuqaÐo diˆnusma z (dhl. mÐa teqnht eikìna), p_{data} h katanom pou akoloujoÔn ta dedomèna eisìdou (stic eikìnec aut ja eÐnai mia polÔ uyhl c diastasimìthtac katanom pou mìno èmmesa kai proseggistikˆ mporeÐ na montelopoi sei èna paragwgikì montèlo tÔpou GAN) kai p_{prior} h prior katanom apì thn opoÐa deigmatoleiptoÔme gia na pˆroume to tuqaÐo diˆnusma sthn eÐsodo tou Generator. Efìson, ìpwc anafèrjhke, o Discriminator bgˆzei pijanìthta kai ˆra $D(x) \in [0,1]$, prokÔptei ìti gia thn elaqistopoiÐhsh thc sunˆrthshc kìstouc tou, o Discriminator prèpei na mˆjei na anajètei uyhl pijanìthta sta deÐgmata me thn etikèta 1 (ta opoÐa proèrqontai apì to sÔnolo dedomènwn ekpaÐdeushc) kai qamhl se autˆ pou parˆgontai apì ton Generator.

To dÐktuo tou Generator, me th seirˆ tou, prospajeÐ na {xegelˆsei} autì tou Discriminator ¸ste oi pijanìthtec pou anajètei sta teqnhtˆ deÐgmata sthn èxodì tou na eÐnai uyhlèc. StoqeÔei, dhlad , na megistopoi sei ton deÔtero ìro thc sunˆrthshc kìstouc tou Discriminator - exˆllou mìno autìn ton ìro mporeÐ na ephreˆsei prokeimènou h sunˆrthsh kìstouc tou Discriminator prokeimènou aut na auxhjeÐ. Epomènwc, gia ton Generator ja isqÔei:

$$J_G(\tilde{\theta}_G, \tilde{\theta}_D) = \frac{1}{m}\sum_{i=1}^{m}\left[(1-y^{(i)})\log(1-h(x^{(i)};\tilde{\theta}))\right] \quad (3.6)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[\log(1-D(G(z^{(i)};\tilde{\theta}_G);\tilde{\theta}_D))\right] \quad (3.7)$$

$$= \frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^{(i)};\tilde{\theta}_G);\tilde{\theta}_D)) \quad (3.8)$$

$$E_{z\sim p_{prior}}\log[1-D(G(z))] \quad (3.9)$$

ìpou to arnhtikì prìshmo sthn arq èqei plèon afairejeÐ kaj¸c o Generator prospajeÐ elaqistopoi¸ntac th dik tou sunˆrthsh kìstouc na aux sei aut tou Discriminator,

en¸ ìla ta upìloipa megèjh eÐnai ìpwc prin. Epeid , o pr¸toc ìroc thc exÐswshc 3.5 exartˆtai mìno apì to sÔnolo dedomènwn ekpaÐdeushc, sunhjÐzetai sth bibliografÐa h parapˆnw sunˆrthsh kìstouc tou Generator na dhl¸netai wc to arnhtikì thc sunˆrthshc kìstouc tou Discriminator:

$$J_G\ \tilde{}_G;\ \tilde{}_C\ =\ J_D\ \tilde{}_C;\ \tilde{}_G \tag{3.10}$$

kˆti pou exhgeÐ giatÐ h ekpaÐdeush enìc GAN qrhsimopoi¸ntac th sunˆrthsh kìstouc Binary Cross-EntropyeÐnai isodÔnamh me èna paÐgnio mhdenikoÔ ajroÐsmatoc dÔo paikt¸n. Tèloc, ìpwc apodeiknÔetai sto [29] autì to st simo (setup) gia ekpaÐdeush twn GANs èqei thn prìsjeth idiìthta ìti h sunˆrthsh kìstouc eÐnai asumptwtikˆ sunep c me th apìstash Jensen-Shannon (Jensen-Shannon divergence - JSD)metaxÔ thc katanom c twn dedomènwn kai aut c pou èqei mˆjei o Generator kai apì thn opoÐa parˆgei ta (teqnhtˆ) deÐgmata sthn èxodì tou.

Çna basikì meionèkthma pou parousiˆzei h qr sh thc Binary Cross-Entropy kai sugkekrimèna h Ôparxh twn logarÐjmwn stic sunart seic kìstouc eÐnai o koresmìc. EmbajÔnontac, ìtan dÔo katanomèc eÐnai {makriˆ} h Binary Cross-EntropypaÐrnei uyhlèc timèc wstìso autèc ftˆnoun se ènan koresmì kai ètsi peraitèrw metakÐnhsh twn katanom¸n den diaforopoieÐ shmantikˆ tic timèc autèc. Kat' epèktash, idiaÐtera sthn arqikˆ fˆsh thc ekpaÐdeushc ìpou h katanom pou èqei mˆjei o Generator apèqei arketˆ apì thn pragmatikˆ poludiˆstath katanom twn dedomènwn ekpaÐdeushc, h Binary Cross-Entropy stic sunart seic kìstouc den dÐnei saf {plhroforÐa} gia to pwc prèpei na metabˆllei tic paramètrouc tou. To Ðdio ìmwc sumbaÐnei kai ìtan autèc oi katanomèc eÐnai sqetikˆ kontˆ, dhlad apì kˆpoio shmeÐo ekpaÐdeushc kai Ôstera. Çtsi, ìpwc faÐnetai kai sto sq ma 25, ìtan oi katanomèc eÐnai arketˆ makriˆ kontˆ to kìstoc parousiˆzei koresmì. Autì prokaleÐ thn exafˆnish twn parag¸gwn (vanishing gradients), kˆti pou èqei odhg sei sth qr sh enallaktik¸n sunart sewn kìstouc gia thn ekpaÐdeush GANs.

Autì to prìblhma eÐnai gnwstì wc vanishing gradients kai odhgeÐ suqnˆ thn ekpaÐdeush enìc GAN se SurrÐknwsh Rujm¸n [2] (Mode Collapse). H SurrÐknwsh Rujm¸n sta pla-

[2] Rujmìc miac pijanotik c katanom c onomˆzetai mÐa {perioq } aut c me uyhl sugkèntrwsh parathr sewn. Gia parˆdeigma, se mÐa Kanonik katanom , h perioq gÔrw apì th mèsh tim eÐnai o monadikìc rujmìc thc katanom c, en¸ upˆrqoun ˆllec katanomèc me perissìterouc apì ènan rujmoÔc qwrÐc anagkastikˆ autoÐ na antistoiqÐzontai se kˆpoia rop . Sto sÔnolo dedomènwn qeirìgrafwn yhfÐwn tou MNIST, gia parˆdeigma, entopÐzoume 10 tètoiouc rujmoÔc, ènan gia kˆje yhfÐo kai h SurrÐknwsh Rujm¸n ja odhgoÔse ènan Generator sthn paragwg eikìnwn apì èna mìno yhfÐo.

Sq ma 25: Grafik apeikìnish tou probl matoc koresmoÔ thc sunˆrthshc kìstouc Binary Cross-Entropy.
Phg : Anakataskeu apì Generative Adversarial Networks Specialization, Zhou et al., DeepLearning.AI, 2021[122]

Ðsia thc Paragwgik c MontelopoÐhshc èqei na kˆnei me thn poikilÐa (diversity) twn deigmˆtwn pou mporeÐ na parˆxei ènac ekpaideumènoc Generator kai sugkekrimèna me th drastik surrÐknwsh aut c se deÐgmata mÐac mìno kathgorÐac klˆshc. —pwc analÔetai apì ton Mao et al. [56], h qr sh thc sunˆrthshc kìstouc Binary Cross-Entropy odhgeÐ ton Generator sto {asfalèc} monopˆti na parˆgei kalˆ deÐgmata mÐac mìno klˆshc, dhlad enjarrÔnei th SurrÐknwsh Rujm¸n. Aut h parat rhsh, èqei afenìc odhg sei sth qr sh èntonhc kanonikopoÐhshc twn sunart sewn kìstouc ekpaÐdeushc GANs, allˆ afetèrou èqei apotelèsei ton lìgo pou plèon den qrhsimopoioÔntai sqedìn poujenˆ sunart seic kìstouc Binary Cross-Entropy sthn ekpaÐdeush twn GANs, me th jèsh touc na èqoun pˆrei ˆllec oi opoÐec epideiknÔoun ligìtero kajìlou koresmì.
Oi basikìterec apì autèc, h sunˆrthsh kìstouc Mèsou TetragwnikoÔ Sfˆlmatoc kai h Apìstash Wasserstein, analÔontai stic paragrˆfouc pou akoloujoÔn.

Sunˆrthsh Kìstouc ElaqÐstwn Tetrag¸nwn (LSGAN)

MÐa ˆllh sunˆrthsh kìstouc pou dokimˆsthke sthn ekpaÐdeush GANs me skopì th stajeropoÐhsh aut c kai th meÐwsh twn peript¸sewn Katˆrreushc Rujm¸n kai exafˆnishc twn parag¸gwn (vanishing gradients), eÐnai h sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn (Least Squares Loss). Gia ton upologismì thc sunˆrthshc kìstouc ElaqÐstwn Tetrag¸nwn qrhsimopoieÐtai to mèso tetragwnikì sfˆlma (mean square error - MSE)metaxÔ twn problèyewn tou Discriminator kai twn {idanik¸n} problèyewn antÐstoiqa gia kˆje dÐktuo kai gia pragmatikèc/teqnhtèc eikìnec (Discriminator: 1 gia tic pragmatikèc eikìnec kai 0 gia tic teqnhtèc, Generator: 1 gia tic teqnhtèc eikìnec pou parˆgei kai me

tic opoÐec jèlei na {xegelˆsei} ton Discriminator ìti eÐnai pragmatikèc).

Prwtoemfanizìmenh sta plaÐsia ekpaÐdeushc GANs sto montèlo LSGAN[56], h sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn basÐzetai sth mèjodo elaqÐstwn tetrag¸nwn apì th Statistik , ìpou prospajoÔme metabˆllontac èna sÔnolo paramètrwn na elaqistopoi soume to ˆjroisma twn tetrag¸nwn twn apostˆsewn sfalmˆtwn miac posìthtac endiafèrontoc apì mÐa posìthta anaforˆc [3]. Sta GANs, ìpwc anafèrjhke, h anaforˆ eÐnai oi ekˆstote etikètec kai h posìthta endiafèrontoc eÐnai oi oi ekˆstote problèyeic tou Discriminator. Se aut thn perÐptwsh, wstìso, o Discriminator den kaleÐtai na bgˆlei sthn èxodì tou mia pijanìthta (dhl. mia tim apì 0 èwc 1, p.q. qrhsimopoi¸ntac Sigmoid sunˆrthsh exìdou), allˆ mia opoiad pote tim , parìlo pou me to pèrac miac epituqhmènhc ekpaÐdeushc aut ja eÐnai kontˆ stic timèc twn etiket¸n.

Antiparabˆllontac me tic exis¸seic 3.5 kai 3.9 thc prohgoÔmenhc paragrˆfou, oi sunart seic kìstouc tic opoÐec prospajoÔn na elaqistopoi soun ta dÐktua enìc GAN pou ekpaideÔetai me th sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn, ja eÐnai:

$$J_D\left(\tilde{\theta}_D;\tilde{\theta}_G\right)=\frac{1}{m}\sum_{i=1}^{m}\left(D\left(x^{(i)};\tilde{\theta}_D\right)-1\right)^2+\left(D\left(G\left(z^{(i)};\tilde{\theta}_G\right);\tilde{\theta}_D\right)-0\right)^2 \tag{3.11}$$

$$=\frac{1}{m}\sum_{i=1}^{m}\left(D\left(x^{(i)};\tilde{\theta}_D\right)-1\right)^2+\left(D\left(G\left(z^{(i)};\tilde{\theta}_G\right);\tilde{\theta}_D\right)\right)^2 \tag{3.12}$$

$$E_{x\sim p_{data}}\left[\left(D\left(x\right)-1\right)^2\right]+E_{z\sim p_{prior}}\left[\left(D\left(G\left(z\right)\right)\right)^2\right] \tag{3.13}$$

gia ton Discriminator, en¸ gia ton Generator antÐstoiqa kai metˆ apì prˆxeic, ja eÐnai

$$J_G\left(\tilde{\theta}_G;\tilde{\theta}_D\right)=\frac{1}{m}\sum_{i=1}^{m}\left(D\left(G\left(z^{(i)};\tilde{\theta}_G\right);\tilde{\theta}_D\right)-1\right)^2 \tag{3.14}$$

$$E_{z\sim p_{prior}}\left[\left(D\left(G\left(z\right)\right)-1\right)^2\right] \tag{3.15}$$

afoÔ o Generator jèlei na {dei} pìso makriˆ apì ta deÐgmata me etikèta 1 katatˆssei o Discriminator ta deÐgmata pou par gage.

Se antÐjesh me th sunˆrthsh kìstouc Binary Cross-Entropy, sto sqhmatismì thc sunˆrthshc kìstouc ElaqÐstwn Tetrag¸nwn den emfanÐzontai oi logˆrijmoi. Autì èqei wc ˆmesh sunèpeia to ˆjroisma twn dÔo ìrwn thc sunˆrthshc kìstouc tou Discriminator

[3] Sun jhc efarmog thc mejìdou ElaqÐstwn Tetrag¸nwn sth Statistik apoteleÐ h Grammik Palindrìmhsh. EkeÐ, dojèntwn kˆpoiwn shmeÐwn dedomènwn, mac zhteÐtai na broÔme mia eujeÐa, oi parˆmetroi thc opoÐac elaqistopoioÔn to ˆjroisma twn tetrag¸nwn twn (kˆjetwn) apostˆsewn apì ta shmeÐa.

na mhn parousiˆzei aut th sigmoeid morf kai ˆra to fainìmeno koresmoÔ ed¸ eÐnai shmantikˆ meiwmèno. Sugkekrimèna, monˆqa sthn perÐptwsh ìpou èxodoc tou Discriminator eÐnai 1 gia ta pragmatikˆ dÐktua kai 0 gia ta teqnhtˆ (dhl. tèleia prìbleyh), h sunˆrthsh kìstouc thc 3.13 ja d¸sei mhdenikˆ gradients[4]. —pwc analÔetai kai sto ˆrjro tou LSGAN, autì èqei san apotèlesma antÐstoiqh meÐwsh se emfˆnish tou fainomènou SurrÐknwshc Rujm¸n se diˆforec arqitektonikèc GANs pou dokimˆsthkan kai shmantikˆ stajerìterh ekpaÐdeush (dhl. pio omalèc timèc twn sunart sewn kìstouc) [56]. Sth pleioyhfÐa twn montèlwn pou anaptÔqjhkan sta plaÐsia thc paroÔsac ergasÐac gÐnetai qr sh thc sunˆrthshc kìstouc ElaqÐstwn Tetrag¸nwn tìso lìgw aut¸n pou elèqjhsan parapˆnw pou èqoun na kˆnoun me thn eustˆjeia thc ekpaÐdeushc, ìso kai lìgw thc taqÔthtac upologismoÔ aut c.

Sunˆrthsh Kìstouc Wasserstein (WGAN)

Tautìqrona sqedìn me th sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn, parousiˆsthke kai mÐa enallaktik sunˆrthsh kìstouc, h opoÐa basÐzetai sthn apìstash MetakÐnhshc EdˆfouC (Earth Mover's Distance - EMD) metaxÔ dÔo katanom¸n kai h opoÐa epÐshc eÐqe san stìqo thn exˆleiyh tou probl matoc thc KatˆrreushcRujm¸n pou suqnˆ sunìdeue thn ekpaÐdeush GANs me sunˆrthsh kìstouc Binary Cross-Entropy. Prìkeitai gia th sunˆrthsh kìstouc Wasserstein, h opoÐa eÐnai mia sunˆrthsh pou metrˆei thn apìstash metaxÔ dÔo katanom¸n kai epeid basÐzetai sthn EMD to kˆnei pio eÔrwsta kai apodotikˆ apì thn Binary Cross-Entropy. Pio sugkekrimèna kai me bˆsh tic dÔo katanomèc pou apeikonÐzontai sto sq ma 25 (mÐa aut pou èqei mˆjei o Generator kai mÐa h katanom twn dedomènwn ekpaÐdeushc), h EMD metrˆei thn apìstash metaxÔ twn dÔo katanom¸n proseggÐzontac to èrgo pou apaiteÐtai gia na gÐnei h paragìmenh katanom (dhl. aut pou èqei mˆjei o Generator) Ðsh me me thn pragmatik . Diaisjhtikˆ, èan h paragìmenh katanom tan swrìc apì q¸ma, h EMD ja mac èdine mia prosèggish thc

[4]—tan anafèroume thn èkfrash { h sunˆrthsh kìstouc ja d¸sei parag¸gouc gradients} ennooÔme to pr¸to b ma tou algorÐjmou back-propagation gia upologismì twn merik¸n parag¸gwn thc sunˆrthshc kìstouc wc proc kˆje ekpaideÔsimh parˆmetro. H èkfrash aut èqkeitai ston upologismì twn parag¸gwn thc sunˆrthshc kìstouc wc proc tic exìdouc tou Discriminator, oi opoÐec dioqeteÔontai arqikˆ mèqri thn eÐsodo tou Discriminator kai katìpin anadromikˆ ftˆnoun stic str¸seic èwc kai thn eÐsodo tou Generator. Koresmìc thc sunˆrthshc kìstouc, epomènwc, eÐnai ìtan to feedback pou lambˆnei o Discriminator sthn èxodì tou den metabˆlletai kajìlou metabˆlletai lÐgo gia mikrèc allagèc thc exìdou tou - kai tìte den mporeÐ na gnwrÐzei proc poia kateÔjunsh prèpei na {kinhjeÐ} ston q¸ro twn ekpaideÔsimwn paramètrwn prokeimènou na mei¸sei th sunˆrthsh kìstouc.

{duskolÐac} metakÐnhshc tou q¸matoc ¸ste o telikìc swrìc (metˆ th metakÐnhsh) na èqei th morf kai na brÐsketai sthn Ðdia jèsh me autìn thc pragmatik c katanom c.

To prìblhma me ta vanishing gradients kai thn Binary Cross-Entropy eÐnai ìti ìso pio makriˆ (ant. kontˆ) eÐnai oi dÔo katanomèc tìso oi timèc teÐnoun ìlo kai pio kontˆ sto 1 (ant. sto 0). AntÐjeta, sthn EMD kai ˆra sth sunˆrthsh kìstouc Wasserstein den upˆrqei kˆpoia mègisth elˆqisth tim - oi timèc pou mporeÐ na pˆrei h sunˆrthsh eÐnai praktikˆ ìlo to R. Çtsi, akìma kai ìtan oi katanomèc eÐnai arketˆ makriˆ, ìpwc sthn arq thc ekpaÐdeushc, oi timèc pou dÐnei hEMD den parousiˆzoun koresmì. —pwc faÐnetai kai sto sq ma 26 pou akolouje Ð, ìtan oi katanomèc eÐnai arketˆ makriˆ kontˆ to kìstoc den parousiˆzei koresmì, kˆti pou diaforopoieÐ thn EMD apì thn Binary Cross-Entropy (bl. sq ma 25). Wc apotèlesma, h qr sh thc EMD stic sunart seic kìstouc ekpaÐdeushc GANs exaleÐfei to fainìmeno exafˆnishc twn parag¸gwn (vanishing gradients), kˆti pou èqei odhg sei sth uiojèths thc (mèsw thc proseggistik c thc morf c, thc sunˆrthshc kìstouc Wasserstein), gia thn ekpaÐdeush GANs. Wc apotèlesma autoÔ, h ekpaÐdeush GANs me th sunˆrthsh kìstouc Wasserstein mei¸nei th pijanìthta katˆrreushc rujm¸n ston Generator, odhg¸ntac ton se pio eustaj ekpaÐdeush pijanìn me kalÔtera apotelèsmata.

Sq ma 26: Grafik apeikìnish thc exìdou thc Apìstashc MetakÐnhshc Edˆfouc (EMD) metaxÔ dÔo katanom¸n, aut c pou èqei mˆjei o Generator kai thc pragmatik c.
Phg : Anakataskeu apì Generative Adversarial Networks Specialization, Zhou et al., DeepLearning.AI, 2021[122]

Epanerqìmenoi sth sunˆrthsh kìstouc Wasserstein (Wasserstein Loss - W-Loss), aut ìpwc anafèrjhke proseggÐzei thn apìstash MetakÐnhshc Edˆfouc (EMD). H prosèggish sta plaÐsia thc ekpaÐdeushc GANs, sÔmfwna me ton Arjovsky et al. sto ˆrjro touc

Wasserstein GAN[69], eÐnai h akìloujh:

$$W\left(P_{data}, P_{gen}\right) = E_{x \sim p_{data}}[f(x)] - E_{z \sim p_{prior}}[f(\hat{x} = G(z))] \tag{3.16}$$

ìpou P_{data}, P_{gen} eÐnai oi katanomèc twn dedomènwn kai tou Generator antÐstoiqa, p_{prior} eÐnai h katanom apì thn opoÐa deigmatolhpteÐtai to tuqaÐo diˆnusma sthn eÐsodo tou Generator kai f eÐnai sunˆrthsh suneq c katˆ Lipschitz[5] (me stajerˆ $K = 1$). O Discriminator (pou ìtan qrhsimopoieÐtai h W-Loss onomˆzetai Critic epeid plèon den kaleÐtai na kˆnei duadik taxinìmhsh - gia lìgouc aplìthtac ja krat soume thn arqik onomasÐa) jèlei na megistopoi sei aut th sunˆrthsh kìstouc apomakrÔnontac katˆ to dunatì perissìtera tic dÔo katanomèc ¸ste na eÐnai pio eÔkolo to èrgo tou, en¸ o Generator jèlei na thn elaqistopoi sei, kaj¸c ètsi ta deÐgmata pou parˆgei ja eÐnai ìlo kai pio kontˆ se autˆ tou sunìlou ekpaÐdeushc. Epomènwc, oi sunart seic kìstouc Wasserstein twn dÔo diktÔwn ja eÐnai:

$$J_C\left(\tilde{\theta}_C; \tilde{\theta}_G\right) = \frac{1}{m}\sum_{i=1}^{m} C\left(x^{(i)}; \tilde{\theta}_C\right) + \frac{1}{m}\sum_{i=1}^{m} C\left(G\left(z^{(i)}; \tilde{\theta}_G\right); \tilde{\theta}_C\right) \tag{3.17}$$

$$E_{x \sim p_{data}}[C(x)] + E_{z \sim p_{prior}}[C(G(z))] \tag{3.18}$$

gia ton Discriminator (sumb. me {c} apì to Critic pou ìpwc anafèrjhke qrhsimopoi jhke sto WGAN), en¸ gia ton Generator antÐstoiqa ja eÐnai:

$$J_G\left(\tilde{\theta}_G; \tilde{\theta}_c\right) = \frac{1}{m}\sum_{i=1}^{m} C\left(G\left(z^{(i)}; \tilde{\theta}_G\right); \tilde{\theta}_C\right) \tag{3.19}$$

$$E_{z \sim p_{prior}}[C(G(z))] \tag{3.20}$$

ìpote kai faÐnetai ìti elaqistopoi¸ntac th sunˆrthsh kìstouc tou o Generator megistopoieÐ ton deÔtero ìro kai ˆra th sunˆrthsh kìstouc tou Discriminator - exoÔ kai h lèxh { adversary} (= antiparˆjesh).

Estiˆzontac sth suneq katˆ 1- Lipschitz sunˆrthsh f, aut sta GANs ja eÐnai to Ðdio to dÐktuo tou Discriminator, to opoÐo lambˆnontac mia eikìna, x, kaleÐtai na d¸sei ènan

[5] MÐa monodiˆstath sunˆrthsh eÐnai suneq c katˆ Lipschitz me stajerˆ K ìtan upˆrqei diplìc k¸noc (se sq ma Q) pou sqhmatÐzetai apì tic eujeÐec $y = Kx$ kai tou opoÐou to kèntro mporeÐ na kineÐtai epˆnw sth sunˆrthsh ètsi ¸ste olìklhrh h kampÔlh thc sunˆrthshc na paramènei pˆnta èxw apì ton diplì k¸no (dhl. mèsa sthn arister kai dexiˆ meriˆ tou Q). Autì exasfalÐzei, ìti se kanèna thc shmeÐo, h sunˆrthsh den auxˆnei pio gr gora apì K. Gia $K = 1$, gia parˆdeigma, o k¸noc apoteleÐtai apì tic eujeÐec $y = x$ kai h sunˆrthsh den mporeÐ se kanèna shmeÐo tou pedÐou orismoÔ thc na auxˆnei pio gr gora apì grammikˆ. Çtsi, sunart seic ìpwc oi ekjetikèc den eÐnai suneq c katˆ Lipschitz me stajerˆ $K = 1$.

pragmatikì arijmì. Epomènwc, h upì suz thsh sunˆrthsh, h opoÐa ja sumbolÐzetai me $c(x)$ (antÐ gia $D(x)$ ìpwc stic prohgoÔmenec paragrˆfouc), ja eÐnai:

$$c : X ! \quad R; \|c\|_L \quad 1 \tag{3.21}$$

ìpou X eÐnai to pedÐo orismoÔ twn eikìnwn tou sunìlou ekpaÐdeushc (p.q. gia ègqrwmec eikìnec $64 Ö 64$ me 8bits/pixel ja eÐnai o q¸roc $255^{3\ 64\ 64} = 255^{12288}$), en¸ h sunj kh sta dexiˆ dhl¸nei ìti h sunˆrthsh prèpei na ikanopoieÐ th sunj kh sunèqeiac 1- Lipschitz. Gia na proseggÐzei epituq¸c èna neurwnikì dÐktuo me ekpaideÔsimec paramètrouc ~ mÐa sunˆrthsh suneq katˆ 1- Lipschitz, ja prèpei h to mètro twn merik¸n tou parag¸gwn thc exìdou tou diktÔou wc proc tic ekpaideÔsimec paramètrouc na eÐnai to polÔ 1 se kˆje shmeÐo tou pedÐou orismoÔ[69]. Çtsi, ìtan èna GAN ekpaideÔetai me sunˆrthsh kìstouc Wasserstein, to neurwnikì dÐktuo tou Discriminator ja prèpei na ikanopoieÐ thn akìloujh sunj kh sunèqeiac ¸ste na eÐnai mia suneq c sunˆrthsh katˆ 1- Lipschitz:

$$r_{\tilde{C}} C\ x; \tilde{C}_{\ 2} \quad 1 \quad 8x\ 2\ X\ () \quad \|c\|_L \quad 1 \tag{3.22}$$

H epibol aut thc sunj khc exasfalÐzei ìti h sunˆrthsh kìstouc eÐnai èqkurh katˆ th mètrhsh thc apìstashc katanom¸n me thn EMD (sthn opoÐa basÐzetai h Wasserstein), me thn ènnoia ìti den eÐnai mìno suneq c kai diaforÐsimh, allˆ kai den auxˆnei uperbolikˆ gr gora .

Praktikˆ, wstìso, h parapˆnw sunj kh eÐnai adÔnato na epiblhjeÐ, diìti apaiteÐ axiolìghsh twn parag¸gwn tou Discriminator se kˆje shmeÐo tou pedÐou orismoÔ tou, dhlad se kˆje pijan eikìna eisìdou. EpÐshc h axiolìghsh ja prèpei na epanalambˆnetai kˆje forˆ pou metabˆllontai oi parˆmetroi tou diktÔou. Çqoun protajeÐ diˆforoi trìpoi gia epibol thc sunj khc thc 3.22 sto neurwnikì dÐktuo tou Discriminator, me kurÐarqec ton Yalidismì Bar¸n (Weight Clipping), thn Poin Parag¸gwn (Gradient Penalty) kai thn KanonikopoÐhsh Fˆsmatoc (Spectral Normalization), oi opoÐec kai analÔontai akoloÔjwc. Sto shmeÐo autì eÐnai skìpimo anaferjeÐ ìti o lìgoc pou proqwrˆme se mia leptomer anˆlush thc sunˆrthshc kìstouc Wasserstein eÐnai ìti brèjhke kai sta dikˆ mac peirˆmata (bl. 6) ìti aut odhgeÐ se pio stajer ekpaÐdeush kai pio realistikèc kai uyhl c poiìthtac eikìnec sta montèla pou ekpaideÔthkan.

Epibol sunj khc sunèqeiac katˆ Lipschitz me Yalidismì Bar¸n

H mèjodoc pou protˆjhke sto WGAN [69] prokeimènou na epiblhjeÐ h sunj kh gia sunèqeia katˆ 1- Lipschitz eÐnai o Yalidismìc twn Bar¸n (Weight Clipping) tou Discriminator

se èna prokajorismèno pedÐo tim¸n. Autì pou epitugqˆnetai me autìn ton trìpo eÐnai ta bˆrh den metabˆllontai polÔ se kˆje b ma kai ˆra èmmesa exasfalÐzetai ìti h parˆgwgoc thc sunˆrthshc tou Discriminator paramènei elegqìmenh. Çtsi, autì pou praktikˆ dokÐmasan oi suggrafeÐc tou WGAN eÐnai se kˆje b ma tou algorÐjmou beltistopoÐhshc tou Discriminator, metˆ thn ananèwsh twn bar¸n tou (dhl. sqedìn ìlwn twn ekpaideÔsimwn paramètrwn), gÐnontan yalidismìc aut¸n kai ètsi ìsa tan megalÔtera apì th prokajorismènh mègisth mikrìtera apì thn elˆqisth metabˆllontan sthn antÐstoiqh mègisth elˆqisth tim (clipping). Katˆ thn ekpaÐdeush tou WGAN protˆjhke oi timèc autèc na eÐnai 0:01.

Monolìti h mèjodoc aut bo jhse sth stajeropoÐhsh kai thn epituqÐa thc ekpaÐdeushc tou WGAN se diˆfora sÔnola dedomènwn, ìpwc parathr jhke lÐgo argìtera, o yalidismìc ekpaideÔsimwn paramètrwn periorÐzei thn ikanìthta mˆjhshc tou Discriminator kaj¸c ton empodÐzei na {entopÐsei} kalˆ topikˆ elˆqista thc sunˆrthshc kìstouc. Kat' epèktash periorÐzontai kai oi anadrˆseic proc ton Generator ˆra genikìtera h qr sh YalidismoÔ Bar¸n sth sunˆrthsh kìstouc Wasserstein mei¸nei thn apìdosh tou GAN. H parat rhsh aut od ghse sthn uiojèthsh miac pio omal c mejìdou epibol c thc sunj khc 3.22, thc Poin c Parag¸gwn pou analÔetai akoloÔjwc.

Enjˆrrunsh sunj khc sunèqeiac katˆ Lipschitz me Poin Parag¸gwn

LÐgo argìtera apì thn arqik prìtash ekpaÐdeushc GANs me th sunˆrthsh kìstouc Wasserstein, o Gulrajani et al. parousiˆsan sto ˆrjro touc { Improved Training of Wasserstein GANs} [71] mia beltÐwsh thc mejìdou YalidismoÔ Bar¸n prokeimènou na epiblhjeÐ h sunj kh thc 3.22 ston Discriminator. H mèjodìc touc, pou onomˆzetai Poin Parag¸gwn (Gradient Penalty), eÐnai arketˆ pio apl kai omal kaj¸c èqkeitai monˆqa sthn prosj kh enìc ìrou kanonikopoÐhshc (regularizing term) sth sunˆrthsh kìstouc tou Discriminator. Pio sugkekrimèna, autì pou kˆnei o ìroc kanonikopoÐhshc eÐnai na anajètei mÐa poin ìtan h nìrma twn merik¸n parag¸gwn thc exìdou tou Discriminator wc proc thn eÐsodì tou eÐnai megalÔterh apì 1.

Wstìso, sÔmfwna me thn 3.22 ja prèpei na elègxoume tic parag¸gouc autèc gia kˆje eÐsodo tou Discriminator, kˆti praktikˆ adÔnato. AntÐ autoÔ, oi suggrafeÐc tou [71] prìteinan na parjoÔn kˆpoia deÐgmata eikìnwn wc grammikèc parembolèc metaxÔ pragmatik¸n kai teqnht¸n eikìnwn kai me bˆsh autˆ na upologistoÔn oi parˆgwgoi kai h poin. Gia parˆdeigma, antÐ h nìrma twn parag¸gwn na upologisteÐ xeqwristˆ gia

mia eikìna apì to sÔnolo ekpaÐdeushc, x, kai mÐa apì thn èxodo tou Generator, $\hat{x}$, ja upologisteÐ se mÐa mÐxh twn dÔo, $\tilde{x} = \epsilon x + (1 - \epsilon)\hat{x}$. Epomènwc, o ìroc kanonikopoÐhshc gia thn anˆjesh poin c se parag¸gouc nìrmac megalÔterhc thc monˆdac, ja eÐnai:

$$reg_{GP} = \mathbb{E}_{\tilde{x}}\left[(\|\nabla_{\tilde{x}} C(\tilde{x})\|_2 - 1)^2\right] \tag{3.23}$$

kai ˆra oi sunart seic kìstouc tic opoÐec prospajoÔn na elaqistopoi soun ta dÔo neurwnikˆ dÐktua enìc GAN to opoÐo ekpaideÔetai me sunˆrthsh kìstouc Wasserstein kai Gradient Penalty (WGAN+GP), ja eÐnai:

$$J_C(\theta_C, \theta_G) = \frac{1}{m}\sum_{i=1}^{m}\left[C\left(x^{(i)}; \theta_C\right)\right] + \frac{1}{m}\sum_{i=1}^{m}\left[C\left(G\left(z^{(i)}; \theta_G\right); \theta_C\right)\right] + \lambda_{GP}\, reg_{GP} \tag{3.24}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[C\left(x^{(i)}; \theta_C\right)\right] + \frac{1}{m}\sum_{i=1}^{m}\left[C\left(G\left(z^{(i)}; \theta_G\right); \theta_C\right)\right] + \lambda_{GP}\left[\frac{1}{m}\sum_{i=1}^{m}\left(\|\nabla_{\tilde{x}} C\left(\epsilon x + (1-\epsilon)\hat{x}; \theta_C\right)\|_2 - 1\right)^2\right] \tag{3.25}$$

$$\mathbb{E}_{x \sim p_{data}}[C(x)] + \mathbb{E}_{z \sim p_{prior}}[C(G(z))] + \lambda_{GP}\,\mathbb{E}_{\tilde{x} \sim p_{\tilde{x}}}\left[(\|\nabla_{\tilde{x}} C(\tilde{x})\|_2 - 1)^2\right] \tag{3.26}$$

gia ton Discriminator, me λ_{GP} na eÐnai suntelest c barÔthtac tou kanonikopoiht , o opoÐoc sto [71] eÐqe tejeÐ sth stajer tim $\lambda_{GP} = 10.0$. Gia ton Generator den upˆrqei kˆpoia allag kai ˆra kai ed¸ ja eÐnai:

$$J_G(\theta_G; \theta_c) = \frac{1}{m}\sum_{i=1}^{m}\left[C\left(G\left(z^{(i)}; \theta_G\right); \theta_C\right)\right] \tag{3.27}$$

$$-\mathbb{E}_{z \sim p_{prior}}[C(G(z))] \tag{3.28}$$

—pwc faÐnetai, h mèjodoc aut den epibˆllei th sunj kh sunèqeiac katˆ 1- Lipschitz, apl¸c enjarrÔnei tic parag¸gouc na mhn apomakrÔnontai apì th monˆda, swstìtera na mhn metabˆllontai polÔ se sqèsh me th metabol thc eikìnac sthn eÐsodo. Wstìso, aut h mèjodoc èqei apodeiqteÐ sto parapˆnw ˆrjro ìti douleÔei kalˆ kai kurÐwc polÔ kalÔtera apì th mèjodo YalidismoÔ Bar¸n. Basikì meionèkthma thc mejìdouc eÐnai h kajustèrhsh pou eisˆgei o prìsjetoc upologismìc parag¸gwn kai parembol¸n. Tèloc, kai gia lìgouc plhrìthtac, parousiˆzetai sthn epìmenh parˆgrafo mia akìmh mèjodoc gia epibol poin c se apìtomec metabolèc twn paramètrwn tou Discriminator, h opoÐa epÐshc èqei dokimasteÐ katˆ thn ekpaÐdeush twn montèlwn GANs thc paroÔsac ergasÐac.

Enjˆrrunsh sunj khc sunèqeiac katˆ Lipschitz me KanonikopoÐhsh FˆsmatocÂ

H kanonikopoÐhsh fˆsmatoc enìc pÐnaka bar¸n eÐnai kˆti pou eÐqe arqikˆ qrhsimopoihjeÐ apì ton Miyato et al. to 2017 gia aÔxhsh thc ikanìthtac genÐkeushc baji¸n neurwnik¸n diktÔwn [85]. Proqwr¸ntac èna b ma peraitèrw, oi suggrafeÐc efˆrmosan èna qrìno argìtera KanonikopoÐhsh Fˆsmatoc sta bˆrh sugkekrimènwn sunlektik¸n str¸sewn tou Discriminator enìc GAN, prokeimènou na enjarrÔnoun katˆ autìn ton trìpo ton Discriminator na ikanopoieÐ th sunj kh sunèqeiac katˆ Lipschitz kˆpoiac stajerˆc K 1 kai ètsi na stajeropoi soun thn ekpaÐdeus tou [96]. Oi suggrafeÐc melèthsan th sumperiforˆ tou montèlou pou anèptuxan, Spectral-Normalized GAN (SN-GAN), ìqi mìno katˆ thn ekpaÐdeush me sunˆrthsh kìstouc Wasserstein allˆ kai me sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn kai diapÐstwsan pwc h KanonikopoÐhsh Fˆsmatoc ston Discriminator odhgeÐ se pio stajer ekpaÐdeush me kalÔtera apotelèsmata se amfìterec tic peript¸seic - kˆti pou mporoÔme na epibebai¸soume kai emeÐc sta montèla pou ekpaideÔsame kai sta opoÐaègine qr sh thc KanonikopoÐhshc Fˆsmatoc.

Apì majhmatik c skopiˆc, to fˆsma enìc pÐnaka, W , (W), eÐnai to sÔnolo ìlwn twn idiotim¸n tou (eˆn prìkeitai gia tetragwnikì pÐnaka) twn idiˆzouswn tim¸n [6] tou (gia mh-tetragwnikì pÐnaka). Epiprìsjeta, h fasmatik (telestik) nìrma tou pÐnaka W , kW k$_2$, eÐnai h tetragwnik rÐza thc mègisthc idiˆzousac tim c autoÔ, dhlad gia th fasmatik nìrma ja isqÔei [117]:

$$kW k_2 = \max^{np}\_ : \quad 2 \quad W^T W^{o} \tag{3.29}$$

Sta neurwnikˆ dÐktua, o pÐnakac W antiproswpeÔei ènan pÐnaka bar¸n kˆpoiac str¸shc se èna TND. Gia thn efarmog thc Fasmatik c KanonikopoÐhshc prèpei kˆje tim tou pÐnaka na diairejeÐ me th fasmatik tou nìrma. Wc apotèlesma, o fasmatikˆ kanonikopoihmènoc pÐnakac, $\overline{W_{SN}}$, mporeÐ na ekfrasteÐ wc:

$$\overline{W_{SN}} = \frac{W}{kW k_2} \tag{3.30}$$

[6] Oi idiˆzousec timèc enìc pÐnaka prokÔptoun apì thn ParagontopoÐhsh Idiˆzouswn Tim¸n (Singular Value Decomposition - SVD). H SVD eÐnai mia genÐkeush thc ParagontopoÐhshc Idiotim¸n, dhlad thc graf c enìc pÐnaka wc ginìmeno pinˆkwn pou perièqoun ta idiodianÔsmata kai tic idiotimèc tou (kanonik morf). H genÐkeush èqkeitai sto ìti h SVD pr¸ta {tetragwnopoieÐ} ènan pÐnaka me pollaplasiasmì me ton anˆstrofì tou kai katìpin brÐskei ta idiodianÔsmata kai tic idiotimèc tou ginomènou - autˆ lègontai idiˆzonta dianÔsmata kai idiˆzousec timèc antÐstoiqa. Çtsi h SVD gia ènan pÐnaka W , ja eÐnai: $W = U\ V^T$, ìpou $U; V$ orjokanonikoÐ pÐnakec kai diag¸nioc pÐnakac me tic idiˆzousec timèc.

Sthn prˆxh, o upologismìc thc SVD tou pÐnaka W, prokeimènou na upologisteÐ h fasmatik tou nìrma, eÐnai arketˆ upologistikˆ apaithtik diadikasÐa, kai gi' autì oi suggrafeÐc tou SN-GAN [96] dokÐmasan mia proseggistik lÔsh: prosèggisan th mègisth idiˆzousa tim kai ta aristerˆ kai dexiˆ idiˆzonta dianÔsmata thc SVD, ũ kai ṽ antÐstoiqa, me mÐa parallag thc epanalhptik c mejìdou dÔnamhc [7]. Pio sugkekrimèna, prìteinan thn epanalhptik prosèggish thc fasmatik nìrmac enìc pÐnaka bar¸n, W, wc ex c:

$$\tilde{u} \; B \; \frac{W^T \tilde{u}}{kW^T \tilde{u} k_2} \tag{3.31}$$

$$\tilde{v} \; B \; \frac{W^T \tilde{v}}{kW^T \tilde{v} k_2} \tag{3.32}$$

$$kW k_2 \quad \tilde{u}^T W \tilde{v} \tag{3.33}$$

ìpou k k$_2$ eÐnai hL2 (EukleÐdeia) nìrma enìc dianÔsmatoc, en¸ oi suggrafeÐc br kan{ pwc mÐa epanˆlhyh arkeÐ gia ikanopoihtik prosèggish } thc fasmatik c nìrmac.

Autì pou apodeiknÔetai sthn anˆlush twn suggrafèwn sto [96] eÐnai ìti h efarmog thc Fasmatik c KanonikopoÐhshc stic teleutaÐec suneliktikèc str¸seic tou Discriminator isodunameÐ me thn epibol poin c sta pr¸ta idiˆzousa stoiqeÐa (singular components) tou pÐnaka bar¸n me prosarmozìmeno suntelest , o opoÐoc apotrèpei to sthloq¸ro tou pÐnaka W apì to na epikentr¸netai se mÐa sugkekrimènh kateÔjunsh metabol c (mèsw twn algorÐjmwn beltistopoÐhshc pou sta GANs sun jwc {koitoÔn} mìno thn pr¸th parˆgwgo) katˆ th diˆrkeia thc ekpaÐdeushc. Autì èqei odhg sei se pio stajer kai eÔrwsth ekpaÐdeush eidikˆ twn pio polÔplokwn montèlwn, en¸ stic perissìterec peript¸seic od ghse kai se kalÔtera paragìmena apotelèsmata. —pwc gÐnetai antilh- ptì apì tic paragìmenec eikìnec sto sq ma 27, h Fasmatik KanonikopoÐhsh uperèqei se ìti aforˆ thn eustˆjeia thc ekpaÐdeushc gia poikÐlec rujmÐseic kai parallagèc twn montèlwn. Tèloc, gia lìgouc plhrìthtac eÐnai skìpimo na anaferjeÐ ìti parìlo pou kˆpoiec sunart seic kìstouc kˆpoiec teqnikèc, ìpwc aut , èqoun kalÔterh jewrhtik upost rixh kai idiìthtec apì kˆpoiec ˆllec, sthn prˆxh èqei apodeiqteÐ ìti diaforetikèc sunart seic kìstouc (kai kanonikopoi seic aut¸n) apodÐdoun kalÔtera

[7] Sthn arijmhtik anˆlush, h mèjodoc dÔnamhc (power method) qrhsimopoieÐtai gia thn eÔresh thc megalÔterhc idiotim c enìc pÐnaka, èstw A. H mèjodoc dÔnamhc mporeÐ na anaparastajeÐ apì tic sqèseic [26]: $v_{\tilde{p}} = Au_{\tilde{p}}$ kai $u_{\tilde{p}+1} = v_{\tilde{p}}/\max(v_{\tilde{p}})$ gia $p = 0; 1; 2; :::$, $A = W^T W$, v_0 arqik prosèggish tou idiodianÔsmatoc tou A kai u_0 s N 0; I . H prosèggish thc mègisthc idiotim c eÐnai to mègisto stoiqeÐo tou ṽ kai h epanalhptik diadikasÐa termatÐzei ìtan h diaforˆ metaxÔ diadoqik¸n mègistwn tim¸n eÐnai mikr .

Sq ma 27: SÔgkrish twn paragìmenwn eikìnwn èxi (6) parallag¸n (kurÐwc wc proc tic paramètrouc tou algìrijmou beltistopoÐhshc) enìc montèlou GAN gia diaforetikèc teqnikèc kanonikopoÐhshc: sthn pr¸th gramm qrhsimopoieÐtai sunˆrthsh kìstouc Wasserstein kai kanonikopoÐhshc Poin c Parag¸gwn, sth deÔterh qrhsimopoieÐtai sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn me kanonikopoÐhsh Apìsbeshc Bar¸n (Weight Decay) kai sthn trÐth qrhsimopoieÐtai epÐshc sunˆrthsh kìstouc ElaqÐstwn Tetrag¸nwn me Fasmatik KanonikopoÐhsh stic teleutaÐec str¸seic tou Discriminator (h sunˆrthsh kìstouc den allˆzei).
Phg : Spectral Normalization for Generative Adversarial Networks , Miyato et al., 2018 [96]

qeirìtera metaxÔ diaforetik¸n montèlwn, allˆ akìmh kai sto Ðdio montèlo metaxÔ diaforetik¸n sunìlwn dedomènwn ekpaÐdeushc.

3.2 Upo-sunj kh Paragwg kai Elegximìthta

Ta montèla ta opoÐa suzht jhkan èwc t¸ra tan sthn pleionìthtˆ touc montèla GANs ta opoÐa prospajoÔsan apl¸c na mimhjoÔn eikìnec apì to sÔnolo ekpaÐdeushc, dhlad ekpaideÔontan qwrÐc epÐbleyh (unsupervised training). Stic dÔo upoenìthtec pou akoloujoÔn, ja parousiˆsoume montèla ta opoÐa ekpaideÔontai me epÐbleyh (supervised training) prokeimènou na mˆjoun desmeumènec pijanotikèc katanomèc, me th sunj kh na eÐnai ˆllote h tˆxh pou jèloume na an kei mia paragìmenh eikìna kai ˆllote sugkekrimèna (optikˆ) qarakthristikˆ pou jèloume aut na èqei. Skopìc aut¸n twn upoenot twn eÐnai na fèroun ton anagn¸sth èna b ma pio kontˆ sta GANs tou pragmatikoÔ kìsmou, ìpou stic perissìterec peript¸seic den arkeÐ h dunatìthta paragwg c realistik¸n eikìnwn pou mimoÔntai èna sÔnolo ekpaÐdeushc, allˆ apaiteÐtai autì na gÐnetai elegqìmena.

Sthn paragwg qwrÐc sunj kh, to GAN ekpaideÔetai lambˆnontac tuqaÐo jìrubo sthn eÐsodo tou Generator kai mia omˆda eÐte teqnht¸n pragmatik¸n eikìnwn sthn eÐsodo tou Discriminator. Çtsi, eˆn epijumoÔme thn paragwg realistik¸n eikìnwn apì mia sugkekrimènh tˆxh me sugkekrimèna qarakthristikˆ, prèpei na dokimˆzoume (evaluate) ton Generator me diaforetikèc tuqaÐec eisìdouc èwc ìtou (apì tÔqh) lˆboume mia eikìna sthn èxodì tou me ta epijumhtˆ qarakthristikˆ. Proc ton skopì autì, afier¸noume aut n thn enìthta, ìpou ja analÔsoume teqnikèc tìso gia paragwg eikìnwn apì sugkekrimènh tˆxh ìso kai gia thn elegximìthta sthn paragwg eikìnwn me stìqo thn Ôparxh sugkekrimènwn qarakthristik¸n.

Upo-sunj kh Paragwg Eikìnwn

Xekin¸ntac me thn Upo-sunj kh Paragwg (Conditional Generation), tìso katˆ th fˆsh ekpaÐdeushc ìso kai katˆ th fˆsh thc dokim c o Generator lambˆnei ektìc apì to diˆnusma tuqaÐou jorÔbou kai plhroforÐa gia thn tˆxh thc eikìnac pou kaleÐtai na parˆgei. AntÐstoiqa, o Discriminator ektìc apì mÐa omˆda eikìnwn kai tic etikètec me thn tˆxh pou antistoiqoÔn sth kˆje mÐa. Epomènwc, se aut n thn perÐptwsh apaiteÐtai h Ôparxh epishmasmènou (annotated) sunìlou dedomènwn prokeimènou na gÐnei h ekpaÐdeush tou GAN.

Conditional GAN (CGAN)

H upo-sunj kh paragwg sta plaÐsia twn GANs, xekin¸ntac apì ton Mirza et al. kai to montèlo Conditional GAN (CGAN)[32], epètreye thn paragwg deigmˆtwn apì opoia-d pote tˆxh etikèta twn dedomènwn - efìson aut èqei {ekproswphjeÐ} ikanopoihti-kˆ sto sÔnolo dedomènwn ekpaÐdeushc. O trìpoc pou protˆjhke sto [32] prokeimènou na dÐnetai h plhroforÐa thc tˆxhc/etikètac tìso ston Generator ìso kai ston Discriminator eÐnai qrhsimopoi¸ntac ta legìmena dianÔsmata one-hot.

Çtsi, sthn perÐptwsh tou Generator, mazÐ me to tuqaÐo diˆnusma sthn eÐsodì tou dÐnetai (sunenwmèno) kai èna diˆnusma m kouc Ðsou me ton arijmì twn diaforetik¸n tˆxe-wn/etiket¸n tou sunìlou ekpaÐdeushc me ìla ta stoiqeÐa mhdenikˆ plhn enìc, autoÔ pou antistoiqeÐ sthn tˆxh pou epijumoÔme na an kei h paragìmenh eikìna, pou eÐnai monˆda. O lìgoc pou paramènei kai to diˆnusma tuqaÐou (gkaousianoÔ sun jwc) jorÔbou sthn eÐsodo, eÐnai diìti kat' autìn ton trìpo upˆrqei kˆpoia tuqaiìthta kai ˆra kˆpoia poikilomorfÐa sta deÐgmata pou parˆgontai apì th sugkekrimènh tˆxh. Autì ìmwc pou exasfalÐzei ìti prˆgmati o Generator ekpaideÔetai gia na parˆgei deÐgmata apì th sugkekrimènh tˆxh pou perigrˆfetai sto one-hot diˆnusma sthn eÐsodì tou, eÐnai to ìti antÐstoiqh plhroforÐa gia thn tˆxh dÐnetai kai ston Discriminator, ìpwc apeikonÐzetai sto sq ma 28 parakˆtw. Sta plaÐsia thc Paragwgik c MontelopoÐhshc eikìnwn, h sunj kh thc tˆxhc dÐnetai ston Discriminator wc ènac one-hot pÐnakac, dhlad ènac tris-diˆstatoc pÐnakac Ðdiou plˆtouc kai m kouc me tic pragmatikèc eikìnec allˆ me bˆjoc (arijmì kanali¸n) ìso kai oi diaforetikèc tˆxeic/etikètec tou sunìlou ekpaÐdeushc. —la autˆ ta kanˆlia perièqoun mhdenikèc timèc plhn enìc, autoÔ pou antistoiqeÐ sthn etikèta thc eikìnac eisìdou to opoÐo èqei monˆdec. Çtsi, ìpwc apeikonÐzetai sto sq ma, ston Generator dÐnetai to tuqaÐo diˆnusma, ≠ kaj¸c kai to diˆnusma one-hot thc tˆxhc. AntÐstoiqa ston Discriminator ta kanˆlia (3 gia èggrwmec eikìnec, 1 gia asprìmaurec) thc eikìnac eisìdouc sunen¸ntai me ta one-hot kanˆlia thc tˆxhc. Autì odhgeÐ ton Generator na mˆjei na parˆgei eikìnec pou an koun sthn epijumht tˆxh twn dedomènwn ekpaÐdeushc.

Tèloc, sto sq ma 29, pou akoloujeÐ faÐnontai paragwgèc tou montèlou CGAN gia kˆje mÐa apì tic 10 diaforetikèc etikètec tou sunìlou ekpaÐdeushc qeirìgrafwn yhfÐwn tou MNIST. Stic paraqjeÐsec eikìnec thc kˆje gramm c tou sq matoc, o Generator lˆmbane sthn eÐsodì tou to Ðdio diˆnusma one-hot sunenwmèno me diˆnusma gkaousianoÔ tuqaÐou jorÔbou. Çtsi oi eikìnec thc kˆje gramm c èqoun paraqjeÐ upo-sunj kh thn

Sq ma 28: Trìpoc embÔjishc thc plhroforÐac thc tˆxhc thc etikètac ston Generator kai Discriminator tou Conditional GAN.
Phg : Anakataskeu apì Conditional Generative Adversarial Nets , Mirza et al., 2014 [32]

embÔjish thc etikètac tou antÐstoiqou yhfÐou.

Sq ma 29: Paragwgèc tou montèlou Conditional GAN (CGAN) to opoÐo èqei ekpaideuteÐ sto sÔnolo dedomènwn qeirìgrafwn yhfÐwn tou MNIST.
Phg : Conditional Generative Adversarial Nets , Mirza et al., 2014 [32]

Elègximh Paragwg

MÐa enallaktik mèjodoc elègqou twn paragìmenwn deigmˆtwn enìc GAN eÐnai autìc na gÐnei ligìtero katˆ th diˆrkeia kai kurÐarqa metˆ to pèrac thc ekpaÐdeushc twn Paragwgik¸n Montèlwn, kˆti pou genikˆ onomˆzetai Elègximh Paragwg (Controllable Generation). H paragwg upo-sunj kh kˆnei qr sh twn etiket¸n tou sunìlou dedomènwn ekpaÐdeushc prokeimènou ta paraqjènta deÐgmata na an koun se kˆpoia epijumht tˆxh. H elègximh paragwg apì thn ˆllh, estiˆzei ston èlegqo twn qarakthristik¸n pou eÐnai epijumhtì na brÐskontai sta paraqjènta deÐgmata, kˆti pou gÐnetai akìmh kai metˆ to pèrac thc ekpaÐdeushc enìc GAN. Se èna dh ekpaideumèno GAN, gia parˆdeigma, h elègximh paragwg isodunameÐ me thn eÔresh ekeÐnhc thc eisìdou apì ton lanjˆnonta q¸ro tou Generator - o opoÐoc onomˆzetai kai q¸roc- z (z-space) - pou odhgeÐ sthn paragwg eikìnwn me sugkekrimèna qarakthristikˆ.

H elègximh paragwg sta GANs estiˆzetai sthn tropopoÐhsh sugkekrimènwn optik¸n qarakthristik¸n thc exìdou enìc Generator. Sto sq ma 30 parakˆtw, gia parˆdeigma, oi suggrafeÐc tou ˆrjrou Interpreting the Latent Space of GANs for Semantic Face Editing [108], anaz thsan mèsw enìc prìsjetou neurwnikoÔ diktÔou tic eisìdouc ekeÐnec enìc pro-ekpaideumènou Generator pou allˆzoun se mÐa arqik paragwg orismèna epiju-mhtˆ optikˆ qarakthristikˆ, ìpwc h hlikÐa, to fÔlo, h pìza k.a. Shmantikì shmeÐo upenjÔmishc eÐnai ìti h ekpaÐdeush tou montèlou pou qrhsimopoi jhke (PGGAN - bl. upoenìthta 4.1.2 tou epìmenou kefalaÐou) eÐqe oloklhrwjeÐ kai Ôstera gÐnontan aut h {anaz thsh} ston z-q¸ro thc eisìdou tou Generator. Sugkekrimèna, oi suggrafeÐc ˆllazan touc lanjˆnontec {kwdikoÔc}, ìpwc touc onìmasan, sthn eÐsodo enìc kalˆ ekpaideumènou montèlou GAN. H pr¸th st lh tou parakˆtw sq matoc faÐnetai h arqik sÔnjesh apì to montèlo PGGAN, en¸ kˆje mÐa apì tic ˆllec st lec faÐnontai ta apotelèsmata tou elègqou enìc sugkekrimènou qarakthristikoÔ.

Epomènwc, se antÐjesh me thn Upo-sunj kh Paragwg , h Elègximh den apaiteÐ epish-masmèna sÔnola dedomènwn ekpaÐdeushc. Çtsi, en¸ sthn pr¸th mporoÔme apl¸c na parˆgoume apì èna ekpaideumèno GAN eikìnec apì mÐa tˆxh, sth deÔterh mporoÔme na parˆgoume pou èqoun perissìtero ligìtero apì kˆpoia epijumhtˆ qarakthristikˆ. MÐa prìsjeth diaforopoÐhsh, eÐnai ìti gia thn upo-sunj kh paragwg apaiteÐtai upo-sunj kh ekpaÐdeush, dhlad sunènwsh thc eisìdou twn diktÔwn enìc GAN me dia-nÔsmata pÐnakec pou antiproswpeÔoun th zhtoÔmenh tˆxh, en¸ h elegximìthta thc paragwg c ègkeitai apl¸c ston qeirismì twn dianusmˆtwn ston z-q¸ro.

Sq ma 30: TropopoÐhsh poikÐlwn qarakthristik¸n tou pros¸pou mèsw thc metabol c twn stoiqeÐwn thc eisìdou tou Generator.
Phg : Interpreting the Latent Space of GANs for Semantic Face Editing , Shen et al., 2019 [108]

—so apl kai na faÐnetai wc idèa kai wc teqnik , h Elegximìthta sthn paragwg den eÐnai eÔkolh oÔte dedomènh. Autì, sÔmfwna me touc suggrafeÐc twn [48] kai [108], ofeÐletai se dÔo basikoÔc lìgouc: sth susqètish (correlation) metaxÔ twn optik¸n qarakthristik¸n ston q¸ro twn eikìnwn exìdou kai sto ìti o z-q¸roc eisìdou eÐnai mperdemènoc entangled. Epexhg¸ntac, ìtan diaforetikˆ qarakthristikˆ ston q¸ro exìdou èqoun uyhl susqètish metaxÔ touc, gÐnetai arketˆ dÔskolo na elègxei kaneÐc to èna qwrÐc na ephreˆsei to ˆllo (ìpwc p.q. sthn perÐptwsh tou fÔlou kai thc Ôparxhc geneiˆdac).

Epiprìsjeta, pollèc forèc o z-q¸roc pou majaÐnei ènac noise-to-image Generator[8] eÐnai mperdemènoc entangled, me thn ènnoia ìti diaforetikèc kateujÔnseic (ˆxonec) ston q¸ro autì den antistoiqoÔn se diaforetikˆ qarakthristikˆ ston q¸ro exìdou. Praktikˆ, allˆzontac ta ènа stoiqeÐo tou tuqaÐou dianÔsmatoc enìc mperdemènou z-q¸rou, en¸ ja jèlame sthn èxodo tou ekpaideumènou Generator na allˆzei ènа qarakth-ristikì, allˆzoun perissìtera apì ènа kai mˆlista ìqi mìnon autˆ pou èqoun susqètish

[8] Sta arqikˆ montèla GANs, allˆ kai se kˆpoia apì ta pio sÔgqrona kai exeligmèna, o Generator lambˆnei ènа tuqaÐo diˆnusma sthn eÐsodì tou kai kaleÐtai na parˆxei mia realistik eikìna. AutoÔ tou tÔpou oi Generators allˆ kai genikìtera h Paragwgik MontelopoÐhsh onomˆzetai noise-to-image, se antidiastol me ˆllouc tÔpouc ìpwc p.q. h paragwg metasqhmatismìc eikìnac-se-eikìna (image-to-image transform) sthn opoÐa h eÐsodoc sunj kh ston Generator eÐnai mia eikìna sthn opoÐa kaleÐtai na kˆnei ènа sÔnolo realistik¸n metasqhmatism¸n.