



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Τομέας Μαθηματικών

Panoptic Segmentation with Deep Neural Networks

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικόλα Ιωάννου

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Γεώργιος Σιόλας
Ε.ΔΙ.Π. Ε.Μ.Π.



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Τομέας Μαθηματικών

Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας
Σημάτων

Panoptic Segmentation with Deep Neural Networks

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Νικόλα Ιωάννου

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Γεώργιος Σιόλας
Ε.ΔΙ.Π. Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 1η Ιουλίου, 2025.

(Υπογραφή)

.....
Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Αντώνιος Συμβώνης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2025

.....
ΙΩΑΝΝΟΥ ΝΙΚΟΛΑΣ

*Διπλωματούχος σχολής Εφαρμοσμένων
Μαθηματικών και Φυσικών Επιστημών Ε.Μ.Π.*

© – All rights reserved. Με επιφύλαξη παντός δικαιώματος.
Νικόλας Ιωάννου, 2025.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικούς σκοπούς. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπούς μη κερδοσκοπικούς, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Στην

Abstract

in

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ.Παναγιώτη Τσανάκα για την επίβλεψη της παρούσας διπλωματικής εργασίας, όπως και τον κ.Γεώργιο Σιόλα για την καθοδήγηση και την συνεργασία την οποία είχαμε. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου γιατί χωρίς αυτούς δεν θα μπορούσα να βρίσκομαι στην θέση την οποία βρίσκομαι τώρα.

Contents

List of Figures	xiii
------------------------	-------------

List of Tables	xiv
-----------------------	------------

1 Μαθηματικό υπόβαθρο	1
2 Εισαγωγή	1
2.1 Κατάτμηση εικόνας	1
2.1.1 Σημασιολογική κατάτμηση εικόνας	2
2.1.2 Κατάτμηση παραδειγμάτων	2
2.1.3 Πανοπτική κατάτμηση εικόνας	3
2.2 Σύνολα δεδομένων	3
2.3 Θεωρητικό υπόβαθρο	3

List of Figures

1	Σημασιολογική κατάτμηση εικόνας	2
2	Κατάτμηση παραδειγμάτων	3
3	Πανοπτική κατάτμηση εικόνας	3

List of Tables

1 Μαθηματικό υπόβαθρο

2 Εισαγωγή

Η όραση υπολογιστών αποτελεί πεδίο της επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης που εστιάζει στον εντοπισμό και την κατανόηση αντικειμένων σε ψηφιακές εικόνες και βίντεο από υπολογιστές. Συγκεκριμένα, το πεδίο αυτό επιχειρεί να αναπαράγει αλγοριθμικά την αίσθηση της όρασης. Διαιρείται σε κατηγορίες με κάποιες από τις κυριότερες εξ' αυτών να είναι:

- Εκτίμηση πόζας (Pose Estimation)
- Ανίχνευση αντικειμένων (Object Detection)
- Κατάτμηση εικόνας (Image Segmentation)
- Αναγνώριση προσώπου (Face recognition)

Η όραση υπολογιστών μπορεί να χρησιμοποιηθεί σε μια ευρύα γκάμα περιπτώσεων όπως στην ιατρική για τον εντοπισμό κακοήθων όγκων, στην αυτόνομη οδήγηση για τον εντοπισμό αντικειμένων και στην ρομποτική.

Η ιστορική της πορεία ξεκινά στις αρχές της δεκαετίας του 1950 με την έρευνα να επικεντρώνεται στην αναγνώριση απλών σχημάτων και μοτίβων μέσω ψηφιοποιημένων εικόνων. Ένα από τα πρώτα σημεία αναφοράς του πεδίου ήταν το Summer Vision Project του MIT το 1966, όπου ανατέθηκε σε φοιτητές να αναπτύξουν ένα ολοκληρωμένο σύστημα υπολογιστικής όρασης σε ένα καλοκαίρι. Αυτό βασίζεται στην τότε αντίληψη πως το πρόβλημα της υπολογιστικής όρασης πως αποτελεί ένα σχετικά απλό πρόβλημα. Κατά την δεκαετία του 1970, οι ερευνητές άρχισαν να διαμορφώνουν τα θεωρητικά θεμέλια της υπολογιστικής όρασης, όπως η ανίχνευση ακμών και η ανακατασκευή τρισδιάστατων σκηνών από δύο διαστάσεις. Στη δεκαετία του 1980 η έρευνα μετατοπίστηκε από την θεωρία στην πράξη με την ανάπτυξη τεχνικών που βασίζονταν κυρίως σε γραμμικά φίλτρα και μετασχηματισμούς (Hough και ο Fourier). Η δεκαετία του 1990 χαρακτηρίστηκε με την αύξηση της υπολογιστικής ισχύος σε στροφή σε στατιστικές προσεγγίσεις και χρήση αλγορίθμων μηχανικής μάθησης. Η δεκαετία του 2010

2.1 Κατάτμηση εικόνας

Η κατάτμηση εικόνας αποτελεί τεχνική στην υπολογιστική όραση όπου μια ψηφιακή εικόνα διαιρείται σε διακριτές περιοχές, με σκοπό τον εντοπισμό αντικειμένων στην εικόνα. Πιο συγκεκριμένα, η κατάτμηση εικόνας αποτελεί την διαδικασία κατά την οποία ανάλογα με την περίπτωση μερικά ή όλα τα εικονοστοιχεία της εικόνας ανατίθενται σε κλάσεις οι οποίες καθορίζουν το αντικείμενο. Η κατάτμηση εικόνας διαιρείται σε κατηγορίες με τις πιο κύριες εξ' αυτών να είναι:

- Σημασιολογική κατάτμηση εικόνας
- Κατάτμηση παραδειγμάτων
- Πανοπτική κατάτμηση εικόνας

2.1.1 Σημασιολογική κατάτμηση εικόνας

Η σημασιολογική κατάτμηση εικόνας αποτελεί τεχνική κατά την οποία σε κάθε εικονοστοιχείο ανατίθεται μια κλάση. Στην τεχνική αυτή δεν γίνεται διαχωρισμός των διαφορετικών στοιχείων της ίδιας κλάσης. Για παράδειγμα, εάν σε μια εικόνα υπάρχουν 2 άνθρωποι, ο αλγόριθμος δεν θα τους διαχωρίσει ως διαφορετικά αντικείμενα.

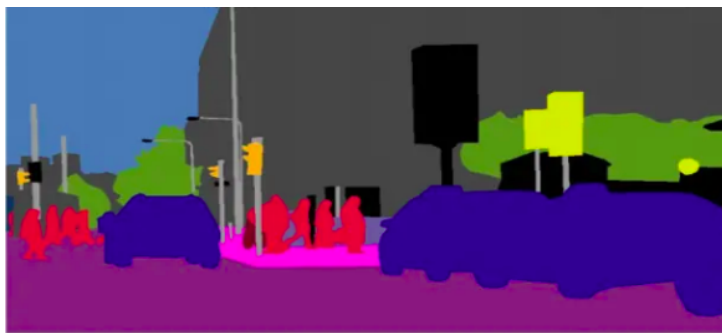


Figure 1: Σημασιολογική κατάτμηση εικόνας

Όπως μπορούμε να δούμε στην εικόνα, τα τρία αυτοκίνητα που παρουσιάζονται αναπαριστώνται με το ίδιο χρώμα, επιβεβαιώνοντας έτσι τα όσα αναφέρθηκαν παραπάνω. Επίσης, μπορούμε να παρατηρήσουμε πως δεν υπάρχει εικονοστοιχείο το οποίο να μην αντιστοιχήστηκε σε κάποια κλάση.

Κάποιες σημαντικές μετρικές απόδοσης της σημασιολογικής κατάτμησης εικόνας είναι η Intersection over Union (IoU), όπως επίσης και η Pixel Accuracy (PA). Μερικά παραδείγματα σημείων αναφοράς (benchmarks) για την μέτρηση της απόδοσης συγκεκριμένων αλγορίθμων που χρησιμοποιούν την τεχνική αυτή είναι τα Cityscapes, PASCAL VOC και ADE20K.

2.1.2 Κατάτμηση παραδειγμάτων

Η κατάτμηση παραδειγμάτων αποτελεί τεχνική κατά την οποία στα μόνο στα εικονοστοιχεία τα οποία αναπαριστούν κάποιο αντικείμενο ανατίθενται κλάση. Στην τεχνική αυτή σε αντίθεση με την σημασιολογική κατάτμηση εικόνας, γίνεται διαχωρισμός των διαφορετικών στοιχείων της ίδιας κλάσης. Για παράδειγμα χρησιμοποιώντας το παράδειγμα που αναφέραμε πριν εάν σε μια εικόνα υπάρχουν περισσότεροι από 1 άνθρωποι, ο αλγόριθμος θα διαχωρίσει κάθε εάν εξ'αυτών ως διαφορετικά αντικείμενα (π.χ. Άνθρωπος 1, Άνθρωπος 2 κ.ο.κ.).

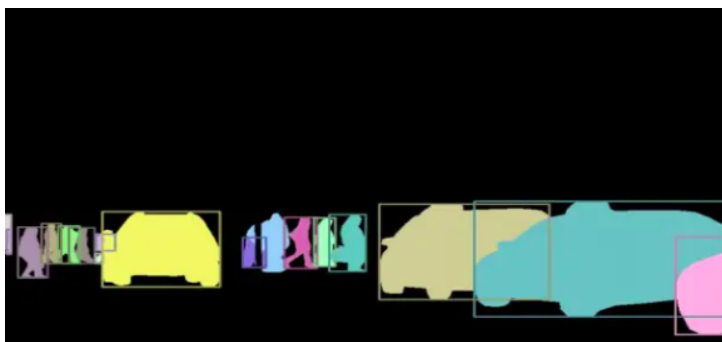


Figure 2: Κατάτμηση παραδειγμάτων

Όπως μπορούμε να δούμε στην εικόνα παραπάνω ο αλγόριθμος κατηγοριοποιεί μόνο τα εικονοστοιχεία τα οποία αντιστοιχούν σε κάποιο αντικείμενο, αδιαφορώντας για τα υπόλοιπα. Εύκολα μπορούμε να παρατηρήσουμε πως κατηγοριοποιούνται μόνο τα αυτοκίνητα και οι άνθρωποι, διακρίνοντας το κάθε αντικείμενο απο το άλλο, ανεξάρτητα εάν ανήκουν στην ίδια κλάση ή όχι.

Κάποιες σημαντικές μετρικές απόδοσης της κατάτμησης παραδείγματος είναι οι Average Precision (AP) και Mask Average Precision (Mask AP). Μερικά παραδείγματα σημείων αναφοράς για την μέτρηση της απόδοσης συγκεκριμένων αλγορίθμων που χρησιμοποιούν την τεχνική αυτή είναι τα COCO, Cityscapes και ADE20K.

2.1.3 Πανοπτική κατάτμηση εικόνας

Η πανοπτική κατάτμηση εικόνας αποτελεί την διαδικασία της ένωσης των δύο προηγούμενων διαδικασιών. Πιο συγκεκριμένα, η τεχνική αυτή



Figure 3: Πανοπτική κατάτμηση εικόνας

2.2 Σύνολα δεδομένων

2.3 Θεωρητικό υπόβαθρο