Κατάτμηση εικόνων με τη μέθοδο ομαδοποίησης UniForCE και χαρακτηριστικά βαθιάς μάθησης

Κωνσταντίνος Ανδρέου

Διπλωματική Εργασία

Επιβλέπων: Αριστείδης Λύκας

Ιωάννινα, Μάρτιος 2024

ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF IOANNINA

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή κ. Λύκα Αριστείδη και τον υποψήφιο διδάκτορα κ. Βαρδάκα Γεώργιο για όλη τη βοήθεια και υποστήριξη που μου παρείχαν κατά τη διάρκεια της εκπόνησης αυτής της διπλωματικής εργασίας.

Ημερομηνία: 24/2/2024

Συγγραφέας: Κωνσταντίνος Ανδρέου

Περίληψη

Στην παρούσα διπλωματική εργασία μελετάται το Βαθύ Συνελικτικό Νευρωνικό Δίκτυο (DCNN) DeepLabV3 που είναι σχεδιασμένο για προβλήματα σημασιολογικής κατάτμησης εικόνας, καθώς και ο αλγόριθμος ομαδοποίησης UniForCE. Εξετάζεται η χρήση τους σε προβλήματα κατάτμησης εικόνας και εστιάζουμε στο πως μπορούμε να χρησιμοποιήσουμε χαρακτηριστικά που εξάγουμε από το προ-εκπαιδευμένο DeepLabV3 για κατάτμηση εικόνας. Αρχικά, ελέγχουμε τις δυνατότητες του UniForCE σε κατάτμηση με ομαδοποίηση με βάση τις τιμές RGB των pixels, ενώ στην συνέχεια εισάγουμε πληροφορία του δικτύου DeepLabV3 για βελτιωμένα αποτελέσματα κατάτμησης. Πιο συγκεκριμένα, χρησιμοποιώντας τις τελικές προβλέψεις του δικτύου DeepLabV3, τα χαρακτηριστικά (features) του επιπέδου ASSP, την χρωματική πληροφορία από το LAB color space και την πληροφορία θέσης των pixel και τροφοδοτώντας τα στον αλγόριθμο ομαδοποίησης UniForCE πετυχαίνουμε πολύ ικανοποιητικά αποτελέσματα κατάτμησης σε εικόνες που δεν περιέχουν αντικείμενα στα οποία έχει εκπαιδευτεί να αναγνωρίζει το δίκτυο DeepLabV3.

Λέξεις Κλειδιά: Συνελικτικά Νευρωνικά Δίκτυα(ΣΝΔ), βαθιά μάθηση, FCN, σημασιολογική κατάτμηση εικόνας, χαρακτηριστικά, προ-εκπαιδευμένο δίκτυο, ομαδοποίηση, PyTorch, DeepLabV3, ASPP, k-means, UniForCE, LAB color space, κατάτμηση.

Abstract

This diploma thesis studies the Deep Convolutional Neural Network (DCNN) DeepLabV3 designed for semantic image segmentation problems, as well as the UniForCE clustering algorithm. Their usage in image segmentation problems is examined and we focus on how we can use features extracted from the pre-trained DeepLabV3 for image segmentation. First, we test the capabilities of UniForCE in segmentation with clustering based on the RGB values of the pixels, while then we introduce information from the DeepLabV3 network for improved results. More specifically, using the final predictions of the DeepLabV3 network, the features of the ASSP layer, the color information from the LAB color space and the pixel position information, and feeding them to the UniForCE clustering algorithm, we achieve highly satisfactory segmentation results in images that do not contain objects that the DeepLabV3 network has been trained to recognize.

Keywords: Convolutional Neural Networks(CNN), deep learning, FCN, semantic segmentation, features, clustering, PyTorch, DeepLabV3, ASPP, k-means, UniForCE, LAB color space, segmentation.

Πίνακας περιεχομένων

Κεφάλαιο 1. Το πρόβλημα της κατάτμησης εικόνων και η συνεισφορά της εργασίας

1.1 Εισαγωγή

Η κατάτμηση εικόνων αποτελεί ένα από τα βασικά προβλήματα στον τομέα της ψηφιακής επεξεργασίας εικόνας και της υπολογιστικής όρασης. Στην ουσία, αναφέρεται στη διαδικασία της διαίρεσης μιας εικόνας σε διακριτές περιοχές, οι οποίες παρουσιάζουν κοινά χαρακτηριστικά. Η κατάτμηση εικόνας έχει πρακτικές εφαρμογές σε πολλούς τομείς όπως η ανάλυση ιατρικών εικόνων, η αναγνώριση προσώπων, η αυτόνομη οδήγηση κ.α. Για την κατάτμηση εικόνας έχουν χρησιμοποιηθεί αρκετές «παραδοσιακές» τεχνικές όπως η κατωφλίωση, η ομαδοποίηση και η ανίχνευση ακμών.

Τα τελευταία χρόνια, με την άνθηση της μηχανικής μάθησης έχει υπάρξει σημαντική βελτίωση στον τομέα της κατάτμησης εικόνας. Η ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων έχει επιτρέψει την επίτευξη προηγουμένως αδύνατων επιδόσεων στα προβλήματα κατάτμησης, με αποτέλεσμα η χρήση τους σήμερα να θεωρείται δεδομένη.

Η παρούσα εργασία εστιάζει στο πρόβλημα της κατάτμησης εικόνας χρησιμοποιώντας πληροφορία από το μοντέλο σημασιολογικής κατάτμησης DeepLabV3[8]. Η σημασιολογική κατάτμηση αποτελεί μια ειδική κατηγορία κατάτμησης εικόνας όπου στόχος της είναι κάθε pixel να κατηγοριοποιηθεί σε μία κλάση, έτσι ώστε pixels της ίδιας κλάσης να αναπαριστούν κάποιο διακριτό αντικείμενο ή μία οντότητα. Τα μοντέλα σημασιολογικής κατάτμησης, ουσιαστικά διαφέρουν από τις άλλες τεχνικές κατάτμησης στο ότι

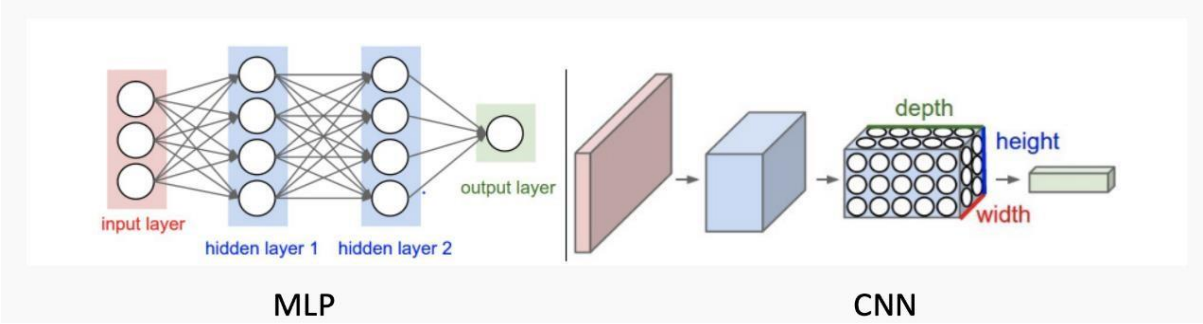
εστιάζουν σε επίπεδο αντικειμένων και όχι σε επίπεδο μεμονωμένων περιοχών. Χρησιμοποιώντας χαρακτηριστικά του μοντέλου DeepLabV3 που έχει εκπαιδευτεί στην αναγνώριση συγκεκριμένων αντικειμένων και τροφοδοτώντας τα στον αλγόριθμο ομαδοποίησης UniForCE[10] προτείνουμε έναν αλγόριθμο κατάτμησης που παρουσιάζει εξαιρετικά αποτελέσματα σε κατάτμηση εικόνων που δεν περιέχουν κατηγορίες αντικειμένων τις οποίες έχει εκπαιδευτεί να αναγνωρίζει το DeepLabV3. Η προσέγγιση μας αυτή προσφέρει ένα επιπλέον επίπεδο λεπτομέρειας και ακρίβειας στην διαδικασία της κατάτμησης, βελτιώνοντας σημαντικά την ποιότητα των αποτελεσμάτων.

Κεφάλαιο 2. Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ)

2.1 Ορισμός

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ)(Convolutional Neural Networks(CNN)) [7] είναι μια κατηγορία αλγορίθμων μηχανικής μάθησης που είναι σχεδιασμένα για να επεξεργάζονται πολυδιάστατα δεδομένα που έχουν τοπολογία πλέγματος, όπως εικόνες και ήχο. Τα ΣΝΔ είναι ένα εξειδικευμένο είδος Νευρωνικών Δικτύων πρόσθιας τροφοδότησης που χρησιμοποιούν συνέλιξη(convolution) στη θέση του γενικού πολλαπλασιασμού πινάκων τουλάχιστον σε ένα από τα επίπεδα(layers) τους.

Παρόλο που η δομή των ΣΝΔ μοιάζει αρκετά με αυτήν των κλασικών Νευρωνικών Δικτύων, διαφέρουν στην οργάνωση των επιπέδων τους. Ένα ΣΝΔ αποτελείται από πολλαπλά επίπεδα(layers), όπου τα τρία κύρια επίπεδα του είναι: το Convolutional Layer, το Pooling Layer και το Fully Connected Layer. Επίσης, οι νευρώνες ενός ΣΝΔ δεν συνδέονται με όλους τους νευρώνες του επόμενου επιπέδου αλλά με μία μικρή περιοχή αυτού, προσφέροντας έτσι αυξημένη ταχύτητα επεξεργασίας των δεδομένων. Τέλος, η έξοδος του δικτύου παρουσιάζεται σε ένα μόνο διάνυσμα που περιέχει τις πιθανότητες της πρόβλεψης κάθε κλάσης.



Εικόνα 1: MLP vs CNN

Αριστερά: φαίνεται ένα κλασικό Νευρωνικό Δίκτυο(MLP) Δεξιά: φαίνεται ένα CNN που έχει σαν είσοδο(φαίνεται με κόκκινο χρώμα) μια εικόνα τριών διαστάσεων(RGB)

Τα ΣΝΔ επιδεικνύουν εξαιρετική αποτελεσματικότητα σε πληθώρα πρακτικών εφαρμογών και χρησιμοποιούνται με μεγάλη επιτυχία στην υπολογιστική όραση. Τα ΣΝΔ χρησιμοποιούνται ευρέως για τον εντοπισμό μοτίβων σε εικόνες, την αναγνώριση αντικειμένων (π.χ. αναγνώριση προσώπων μέσα σε εικόνες), καθώς και την κατηγοριοποίηση των αντικειμένων σε διάφορες κλάσεις.

2.2 Συνέλιξη

Στα μαθηματικά η συνέλιξη είναι μία μαθηματική πράξη δύο συναρτήσεων (f και g), που παράγει μία νέα συνάρτηση ($f * g$). Στην βασικότερη της μορφή, η συνέλιξη αφορά τον τρόπο με τον οποίο ένα σήμα ή μία συνάρτηση μπορεί να επηρεάζεται από ένα άλλο σήμα, συνήθως με τον τρόπο της επικάλυψης των δύο σημάτων. Ο μαθηματικός τύπος της συνέλιξης δίνεται ως εξής:

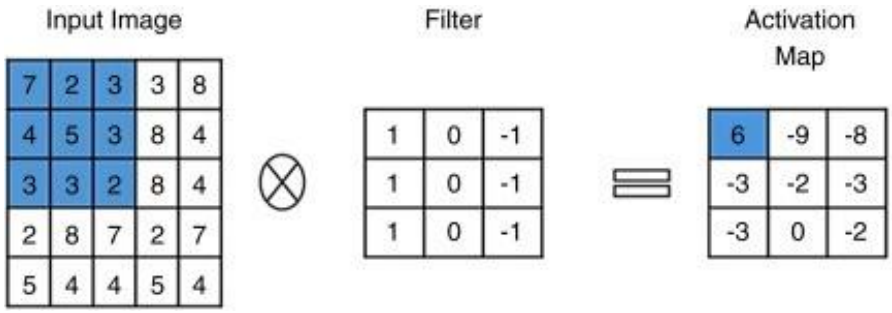
$$s(t) = (x * w)(t) = \int x(\tau)w(t-\tau)d\tau$$

Ωστόσο, στην περίπτωσή μας τα δεδομένα που επεξεργαζόμαστε είναι διακριτά και πεπερασμένα, οπότε χρειάζεται να αλλάξουμε το ολοκλήρωμα και να δουλέψουμε με πεπερασμένα αθροίσματα. Ο τύπος της συνέλιξης που θα χρησιμοποιήσουμε είναι ο εξής:

$$s(t) = (x * w)(t) = \sum x(\tau)w(t-\tau)d\tau.$$

Στα ΣΝΔ[1], η διακριτή συνέλιξη χρησιμοποιείται για την εξαγωγή χαρακτηριστικών από διακριτές εισόδους, όπως εικόνες. Το πρώτο όρισμα της συνέλιξης αναφέρεται ως είσοδος (input) και το δεύτερο ως φίλτρο (filter/kernel). Το αποτέλεσμα συνήθως

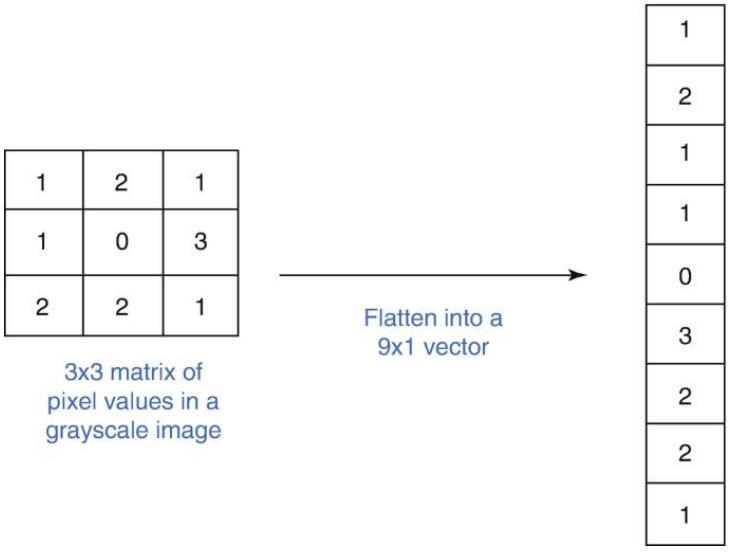
αναφέρεται ως χάρτης χαρακτηριστικών (feature map). Κατά την διαδικασία της διακριτής συνέλιξης, το φίλτρο μετακινείται πάνω από την εικόνα εισόδου και εκτελείται ένας πολλαπλασιασμός πινάκων σε κάθε τοποθεσία της εισόδου με αποτέλεσμα την σύνθεση του χάρτη χαρακτηριστικών.



Εικόνα 2: Παράδειγμα συνέλιξης

2.3 Πλεονεκτήματα CNN έναντι MLP

Η μετατροπή μιας εικόνας σε ένα μονοδιάστατο διάνυσμα και η χρήση της σε ένα MLP για κατηγοριοποίηση θα μπορούσε να λειτουργήσει σε κάποιες απλές δυαδικές εικόνες.



Εικόνα 3: Μετατροπή μίας εικόνας 3x3 σε ένα μονοδιάστατο διάνυσμα 9x1

Ωστόσο, όταν οι εικόνες γίνονται πιο πολύπλοκες και σύνθετες, όπου υπάρχουν χωρικές εξαρτήσεις μεταξύ των εικονοστοιχείων, αυτή η προσέγγιση παρουσιάζει μικρή έως μηδαμινή ακρίβεια.

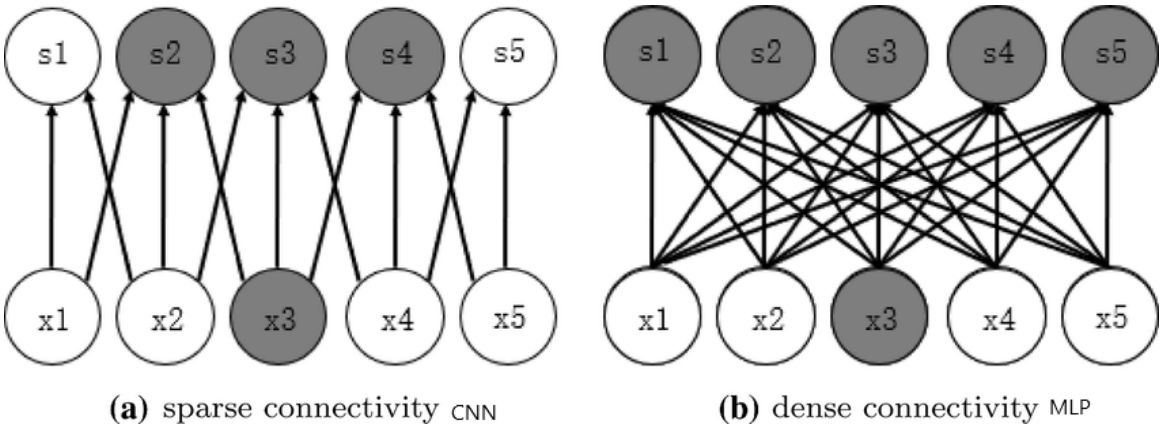
Επιπλέον, η πλήρης συνδεσιμότητα των νευρώνων στα MLP αυξάνει ραγδαία τον αριθμό των βαρών κάνοντας έτσι μη διαχειρίσιμο των αριθμό των παραμέτρων που πρέπει να ρυθμιστούν.

Από την άλλη, ένα CNN είναι σε θέση να αντιμετωπίσει αποτελεσματικά τις χωρικές και προσωρινές εξαρτήσεις σε μια εικόνα μέσω της εφαρμογής κατάλληλων φίλτρων. Αυτό επιτυγχάνεται μέσω των τριών κύριων ιδεών[1] που χρησιμοποιούνται στη συνέλιξη: οι αραιές αλληλεπιδράσεις (sparse interactions ή sparse connectivity), η κοινή χρήση παραμέτρων (parameter sharing) και οι εξισωτικές αναπαραστάσεις (equivariant representations).

Κατά συνέπεια, τα CNN είναι πιο αποτελεσματικά από τα MLP στην ανάλυση και κατανόηση πολύπλοκων εικόνων, επειδή εκμεταλλεύονται αποτελεσματικά τις δομικές πληροφορίες που περιέχονται σε αυτές.

2.3.1 Αραιές αλληλεπιδράσεις (sparse interactions)

Τα κλασικά νευρωνικά δίκτυα χρησιμοποιούν πολλαπλασιασμό με ένα πίνακα παραμέτρων που περιγράφουν την αλληλεπίδραση μεταξύ της μονάδας εισόδου με κάθε μονάδα εξόδου. Αυτό σημαίνει ότι κάθε μονάδα εξόδου αλληλοεπιδρά με κάθε μονάδα εισόδου. Ωστόσο τα νευρωνικά δίκτυα συνέλιξης έχουν αραιή αλληλεπίδραση, λόγω του ότι το φίλτρο συνέλιξης που χρησιμοποιείται έχει μικρότερες διαστάσεις από την είσοδο. Για παράδειγμα, μια εικόνα μπορεί να έχει εκατομμύρια ή χιλιάδες pixels, αλλά κατά την επεξεργασία της χρησιμοποιώντας το φίλτρο μπορούμε να ανιχνεύσουμε σημαντικές πληροφορίες που αφορούν μόνο μερικές εκατοντάδες ή χιλιάδες pixels. Αυτό σημαίνει ότι αποθηκεύουμε λιγότερες παραμέτρους, οι οποίες όχι μόνο μειώνουν τις απαιτήσεις μνήμης του μοντέλου, αλλά βελτιώνουν επίσης τη στατιστική απόδοση του.



Εικόνα 4: α) Αραιές αλληλεπιδράσεις που συναντώνται στα CNN, β)πυκνές αλληλεπιδράσεις που συναντώνται στα MLP

Στην πρώτη περίπτωση του CNN η είσοδος x3 λόγω συνέλιξης της με έναν πίνακα πλάτους τρία επηρεάζει μόνο τρεις εξόδους. Στην δεύτερη του MLP επηρεάζονται όλες οι έξοδοι.

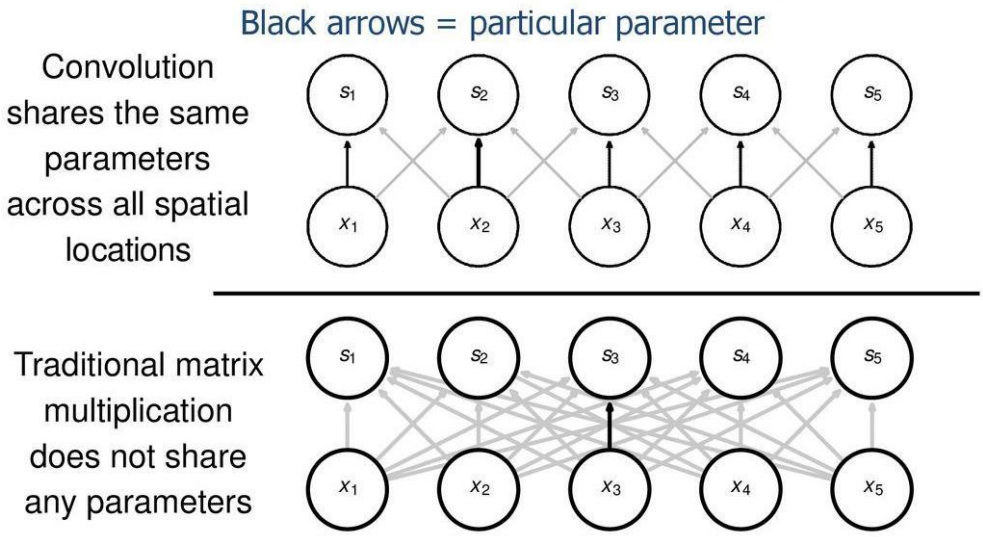
2.3.2 Κοινή χρήση παραμέτρων(parameter sharing)

Η κοινή χρήση παραμέτρων είναι ένα σημαντικό χαρακτηριστικό των Convolutional Neural Networks (CNN). Στα CNN, οι ίδιες παράμετροι ή φίλτρα χρησιμοποιούνται κατά

την εφαρμογή των φίλτρων σε διαφορετικές περιοχές της εικόνας. Αυτό σημαίνει ότι τα ίδια βάρη επαναχρησιμοποιούνται σε διαφορετικά σημεία(εικονοστοιχεία) της εικόνας, αντί να έχουμε ξεχωριστά σετ παραμέτρων για κάθε διαφορετική τοποθεσία.

Αυτή η τεχνική της χρήσης των ίδιων παραμέτρων πολλές φορές οδηγεί στην μείωση του αριθμού των παραμέτρων μειώνοντας την πολυπλοκότητα του μοντέλου και τις απαιτήσεις για υπολογιστικούς πόρους, ενώ ταυτόχρονα βοηθάει στην αντιμετώπιση του προβλήματος της υπερεκπαίδευσης. Επιπλέον, επιτρέπει στο μοντέλο να μάθει πιο γενικευμένα χαρακτηριστικά που είναι χρήσιμα σε διάφορα σημεία της εικόνας, ανεξάρτητα από την τοποθεσία.

Συνολικά, η κοινή χρήση παραμέτρων στα CNN συμβάλλει στη βελτίωση της απόδοσης του μοντέλου, στη μείωση της πολυπλοκότητάς του και στην επίτευξη καλύτερης γενίκευσης σε προβλήματα υπολογιστικής όρασης.



Εικόνα 5: Parameter Sharing(κοινή χρήση παραμέτρων)

Τα μαύρα βέλη αναπαριστούν τις συνδέσεις που αξιοποιούν μια συγκεκριμένη παράμετρο σε δύο διαφορετικά μοντέλα. (Πάνω): Τα μαύρα βέλη δείχνουν τις χρήσεις του κεντρικού στοιχείου ενός πίνακα τριών στοιχείων σε ένα συνελικτικό μοντέλο. Λόγω της κοινής χρήσης παραμέτρων, αυτή η παράμετρος χρησιμοποιείται σε όλες τις θέσεις εισόδου. (Κάτω): Το μοναδικό μαύρο βέλος δείχνει τη χρήση του κεντρικού στοιχείου του πίνακα βαρών σε ένα πλήρως συνδεδεμένο μοντέλο. Αυτό το μοντέλο δεν έχει κοινή χρήση παραμέτρων, οπότε η παράμετρος χρησιμοποιείται μόνο μια φορά.

2.3.3 Εξισωτικές αναπαραστάσεις(equivariant representations)

Στις συνελίξεις, η συγκεκριμένη περίπτωση κοινής χρήσης παραμέτρων οδηγεί τα επίπεδα του δικτύου στο να έχουν μια ιδιότητα που ονομάζεται ισοδυναμία(**equivariance to translation**). Για να πούμε ότι μια συνάρτηση είναι ισοδύναμη σημαίνει ότι εάν αλλάξει η είσοδος, η έξοδος αλλάζει με τον ίδιο τρόπο. Δηλαδή μια συνάρτηση f είναι **ισοδύναμη** με μια συνάρτηση g ανν $f(g(x)) = g(f(x))$.

Κατά την επεξεργασία δεδομένων χρονοσειρών, αυτό σημαίνει ότι η συνέλιξη παράγει ένα είδος χρονοδιαγράμματος για το πότε εμφανίζονται διαφορετικά χαρακτηριστικά στην είσοδο. Εάν μετακινήσουμε ένα γεγονός αργότερα στο χρόνο, η ίδια ακριβώς αναπαράστασή του θα εμφανιστεί στην έξοδο, λίγο αργότερα στο χρόνο.

Ομοίως, για τις εικόνες, η συνέλιξη δημιουργεί έναν δισδιάστατο χάρτη όπου ανιχνεύονται ορισμένα χαρακτηριστικά στην είσοδο. Εάν μετακινήσουμε το αντικείμενο στην είσοδο, θα μετακινηθεί κατά την ίδια ποσότητα στην έξοδο.

Για παράδειγμα, δεδομένης μιας συνάρτησης I, η οποία αναπαριστά τη φωτεινότητα της εικόνας σε ακέραιες τιμές, και μιας άλλης συνάρτησης g, η οποία μετακινεί κάθε εικονοστοιχείο της I προς τα δεξιά κατά μία θέση, δηλαδή $I'=g(I)$ και $I'(x,y)=I(x-1,y)$, η εφαρμογή της g πρώτα στην I και στη συνέχεια η εκτέλεση της συνέλιξης, δίνει το ίδιο αποτέλεσμα σαν να είχε εφαρμοστεί η συνέλιξη στο I' και μετά ο μετασχηματισμός g στο αποτέλεσμα.

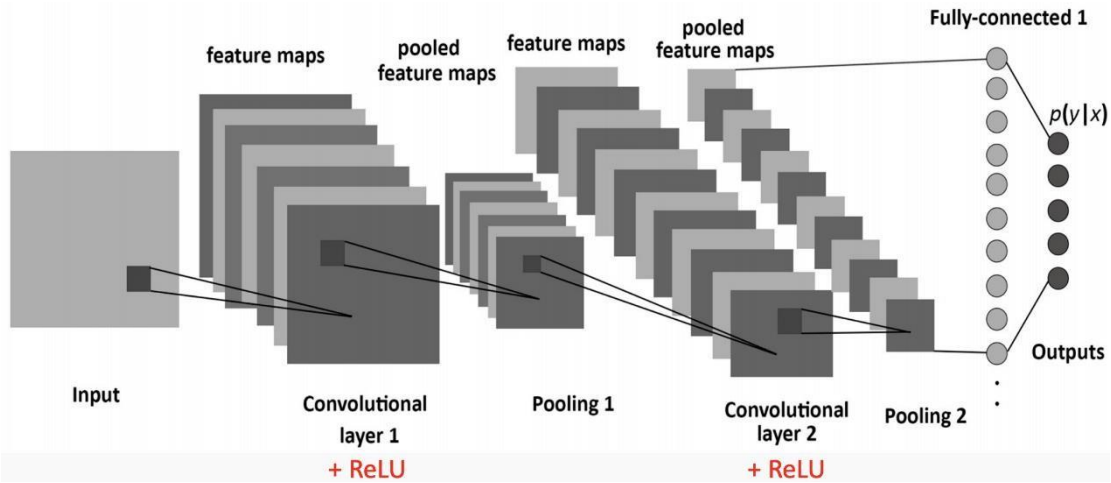
2.4 Επίπεδα(Layers) ενός CNN

Τα επίπεδα ενός τυπικού συνελικτικού νευρωνικού δικτύου(CNN) αποτελούν τα βασικά συστατικά του μοντέλου και επιτελεί το καθένα μία σημαντική λειτουργία για την επεξεργασία των δεδομένων εισόδου. Κάθε επίπεδο λειτουργεί συνεργατικά για να εξάγει χρήσιμα χαρακτηριστικά από τα δεδομένα και να μετατρέψει την είσοδο σε μια πιο ενδελεχή αναπαράσταση που είναι κατάλληλη για την επίλυση του εκάστοτε προβλήματος.

Τα επίπεδα ενός CNN αποτελούνται συνήθως από **Convolutional layers**, **ReLu layers**, **Pooling layers** και ένα **Fully Connected layer**.

Μια τυπική αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου έχει ως εξής:

Είσοδος->Convolution->ReLU->Convolution->ReLu->Pooling->ReLU->Convolution->ReLU->Pooling->Fully Connected.

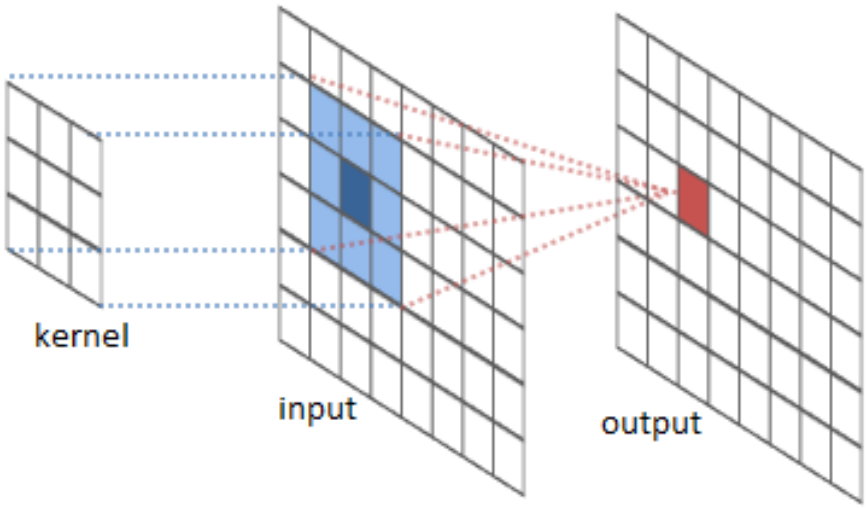


Εικόνα 6: Αρχιτεκτονική ενός CNN που αποτελείται από δυο Convolutional layers ακολουθούμενα από ReLu layers, δύο Pooling layers και ένα Fully Connected layer με 5 εξόδους.

2.4.1 Convolutional Layer

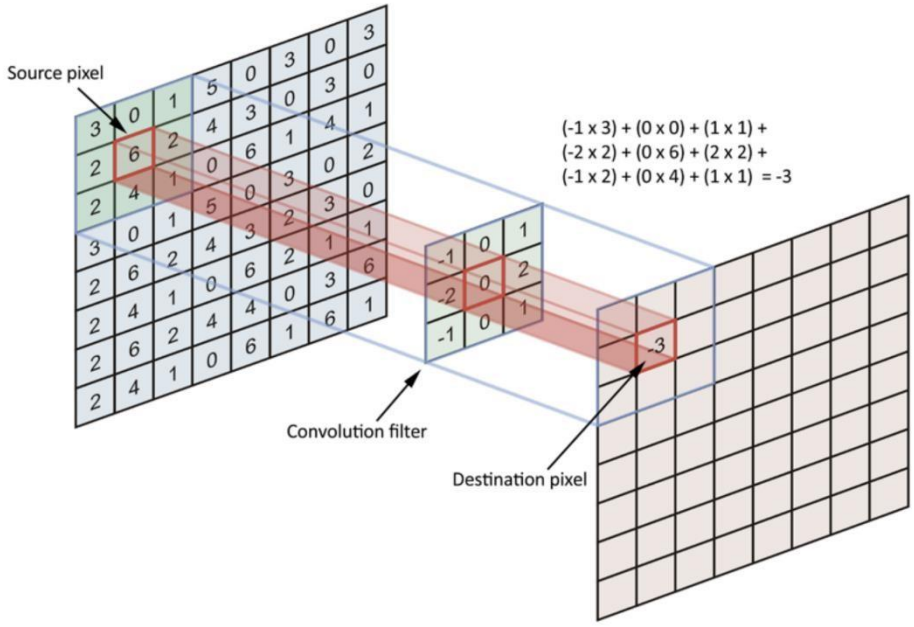
Το επίπεδο συνέλιξης έχει κομβικό ρόλο στη λειτουργία των ΣΝΔ και ο βασικός στόχος της λειτουργίας του είναι η εξαγωγή χαρακτηριστικών και μοτίβων που παρουσιάζονται στην εικόνα εισόδου.

Στο επίπεδο αυτό εκτελείται η συνέλιξη της εικόνας με φίλτρα τα οποία έχουν προκαθοριστεί. Το είδος των φίλτρων που θα χρησιμοποιηθούν και οι διαστάσεις τους διαφέρουν ανάλογα με το πρόβλημα που έχουμε να αντιμετωπίσουμε. Οι συνηθισμένες διαστάσεις των φίλτρων είναι μικρές, π.χ. 3x3, 5x5 αλλά εκτείνονται στο πλήρες βάθος του όγκου εισόδου.



Εικόνα 7: Λειτουργία Επίπεδου Συνέλιξης

Κατά τη διάρκεια του εμπρόσθιου περάσματος(forward pass), κάθε φίλτρο ολισθαίνει κατά το πλάτος και το ύψος της εισόδου και υπολογίζεται το γινόμενο των τιμών του φίλτρου και της εισόδου, παράγοντας την αναπαράσταση κάθε περιοχής στις δύο διαστάσεις, τον γνωστό ως χάρτη ενεργοποίησης για το συγκεκριμένο φίλτρο. Στοιβάζοντας όλους τους χάρτες ενεργοποίησης που προκύπτουν από κάθε φίλτρο(άρα υπάρχουν τόσοι χάρτες όσοι και τα φίλτρα), δημιουργείται ένας κύβος εξόδου, ο οποίος αποτελείται από έναν δισδιάστατο χάρτη ανά φίλτρο. Συνήθως, διαφορετικά φίλτρα ανιχνεύουν διαφορετικά χαρακτηριστικά στην εικόνα εισόδου. Συμβατικά, το πρώτο επίπεδο συνέλιξης είναι υπεύθυνο για τη σύλληψη χαρακτηριστικών χαμηλού επιπέδου, όπως οι ακμές, τα χρώματα και ο προσανατολισμός της κλίσης. Με την προσθήκη πρόσθετων επιπέδων, η αρχιτεκτονική προσαρμόζεται επίσης στα χαρακτηριστικά υψηλού επιπέδου.



Εικόνα 8: Παράδειγμα συνέλιξης ενός φίλτρου

Το μέγεθος του όγκου εξόδου καθορίζεται από τις ακόλουθες βασικές παραμέτρους:

- Το **βήμα(stride) συνέλιξης** καθορίζει την απόσταση την οποία το φίλτρο μετακινείται κατά μήκος της εισόδου κατά τη συνέλιξη. Όταν το stride είναι 1 τότε σημαίνει ότι μετακινούμε το φίλτρο ένα pixel τη φορά. Μειώνοντας το βήμα συνέλιξης, αυξάνεται το μέγεθος του όγκου εξόδου.
- Το **zero-padding** αναφέρεται στα μηδενικά που μπορεί να τοποθετήσουμε κάποιες φορές γύρω από τα όρια της εισόδου πριν την εφαρμογή της συνέλιξης. Η τεχνική αυτή χρησιμοποιείται για τη διατήρηση των χωρικών διαστάσεων της εικόνας εισόδου μετά το πέρας της συνέλιξης σε έναν χάρτη χαρακτηριστικών και την αποφυγή απώλειας πληροφοριών στα όρια της εικόνας. Τέλος, η συμπλήρωση μπορεί να χρησιμοποιηθεί για να ελέγξουμε το μέγεθος

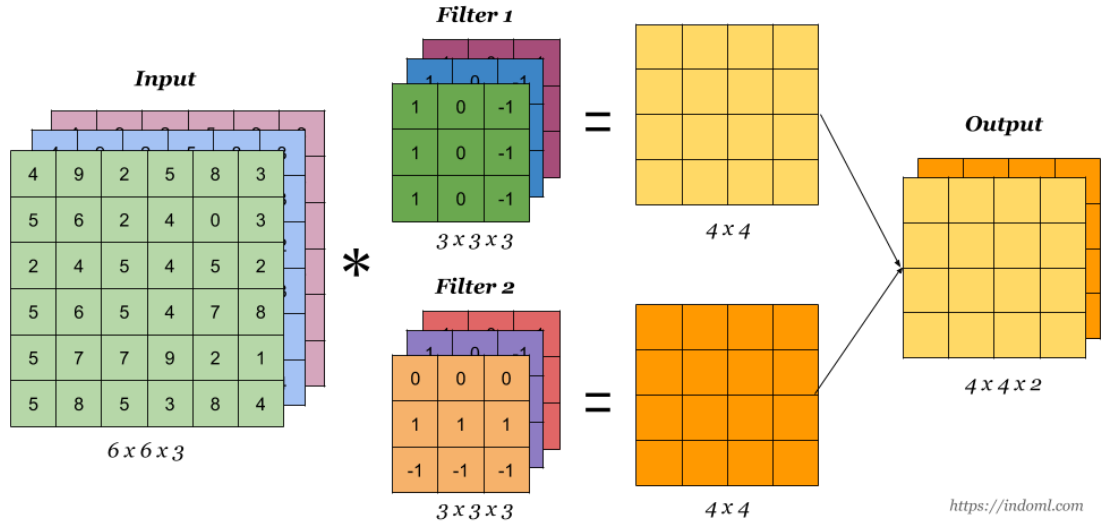
του όγκου εξόδου, ιδίως όταν χρησιμοποιούνται φίλτρα με μεγαλύτερο μέγεθος.

- **Ο αριθμός των φίλτρων** καθορίζει το βάθος της εξόδου.

Αν έχουμε μια εικόνα διάστασης **W**, stride **S**, zero padding **P** και φίλτρο διάστασης **K**, τότε το μέγεθος του όγκου εξόδου των χαρτών ενεργοποίησης δίνεται από τον τύπο:

$$(W - K + 2P) / S + 1,$$

Θέτοντας το zero-padding ίσο με P=(K-1)/2 όταν stride=1 εξασφαλίζουμε ότι ο όγκος εξόδου θα έχει ίδιες χωρικές διαστάσεις με τον όγκο εισόδου.



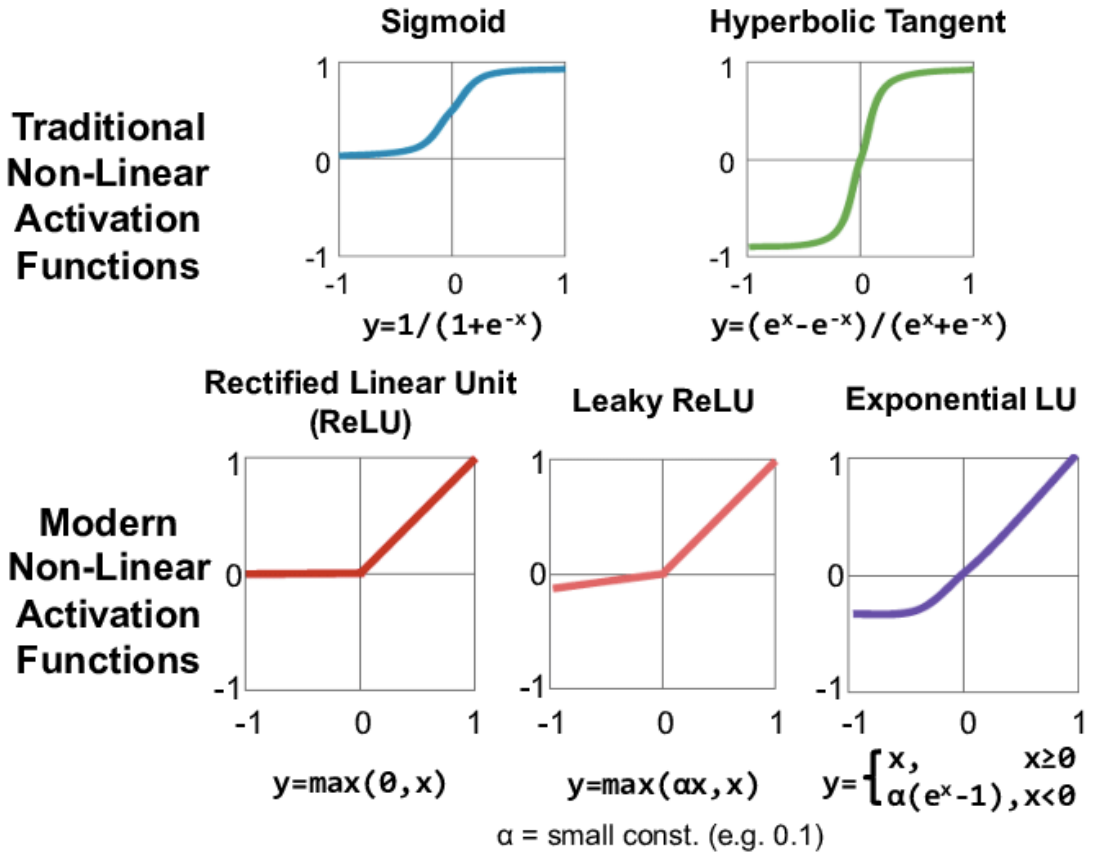
Εικόνα 9: Παράδειγμα συνέλιξης με δύο φίλτρα

2.4.2 ReLU layer

Η συνάρτηση **ReLU**(Rectified Linear Unit) είναι μία πολύ δημοφιλής συνάρτηση ενεργοποίησης στα προβλήματα μηχανικής μάθησης, η οποία υπολογίζει τη συνάρτηση $f(\kappa)=\max(0,\kappa)$. Με άλλα λόγια, η συνάρτηση παίρνει την τιμή μηδέν για αρνητικές εισόδους και για θετικές τιμές επιστρέφει την είσοδο της.

Μετά το Convolutional layer εφαρμόζεται η συνάρτηση ReLU στους χάρτες χαρακτηριστικών, ώστε να αυξηθεί η μη γραμμικότητα στο δίκτυο. Το γεγονός ότι το επίπεδο αυτό τοποθετείται μετά το Convolutional layer οφείλεται στο ότι η συνέλιξη είναι μια γραμμική λειτουργία. Τα πραγματικά δεδομένα που θέλουμε να μάθει ένα CNN θα είναι μη γραμμικά, όπως μη γραμμικές είναι και οι εικόνες, και μπορούμε να στηριχθούμε σε αυτό με μια συνάρτηση ενεργοποίησης όπως η ReLU.

Άλλες συναρτήσεις ενεργοποίησης που μπορούν να χρησιμοποιηθούν για να ενισχύσουμε την μη γραμμικότητα είναι η υπερβολική εφαπτομένη(**tanh**) και η σιγμοειδής συνάρτηση(**sigmoid σ**), αλλά και παραλλαγές της ReLU όπως η **Leaky ReLU.**



Εικόνα 10: Μη γραμμικές συναρτήσεις ενεργοποίησης που μπορούν να χρησιμοποιηθούν σε ένα CNN.

2.4.3 Pooling layer

Τα επίπεδα συγκέντρωσης/υποδειγματοληψίας **(Pooling layer)** συνήθως εισάγονται μετά από επίπεδα συνέλιξης. Ο βασικός στόχος της συγκέντρωσης(pooling) είναι η μείωση του μεγέθους των χαρτών ενεργοποίησης, η οποία με τη σειρά της οδηγεί στην μείωση των παραμέτρων της εκπαίδευσης και στην αύξηση της ταχύτητας επεξεργασίας. Επίσης, το επίπεδο συγκέντρωσης συνοψίζει τα χαρακτηριστικά που υπάρχουν σε μια περιοχή του χάρτη χαρακτηριστικών που δημιουργείται από ένα επίπεδο συνέλιξης. Έτσι, δίνεται η δυνατότητα στο δίκτυο να εντοπίσει αντικείμενα σε μια εικόνα ανεξάρτητα από το σημείο που αυτά βρίσκονται, καθιστώντας το μοντέλο πιο ανθεκτικό στις παραλλαγές στη θέση των χαρακτηριστικών στην εικόνα εισόδου.

2.4.3.1 Λειτουργία του Pooling Layer

Η συγκέντρωση περιλαμβάνει την επιλογή ενός φίλτρου συγκέντρωσης που θα εφαρμοστεί στους χάρτες χαρακτηριστικών. Το μέγεθος του φίλτρου είναι μικρότερο από το μέγεθος του χάρτη χαρακτηριστικών, με συνηθισμένη διάσταση 2×2 και βήμα(stride) 2 pixels.

Αυτό σημαίνει ότι το επίπεδο συγκέντρωσης θα μειώνει πάντα το μέγεθος κάθε χάρτη χαρακτηριστικών κατά 2, π.χ. κάθε διάσταση μειώνεται στο μισό, μειώνοντας τον αριθμό των pixels ή των τιμών σε κάθε χάρτη χαρακτηριστικών στο ένα τέταρτο του αρχικού μεγέθους. Για παράδειγμα, ένα επίπεδο συγκέντρωσης που εφαρμόζεται σε έναν

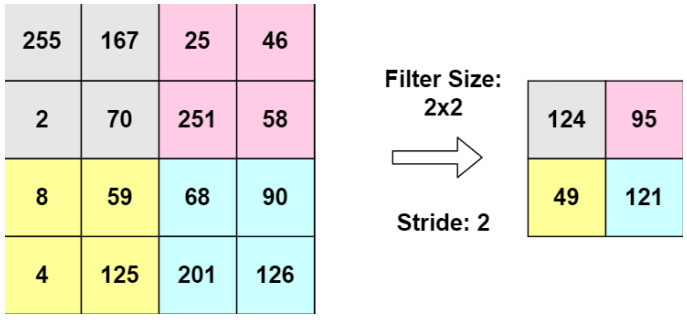
χάρτη χαρακτηριστικών 6×6 (36 εικονοστοιχεία) θα έχει ως αποτέλεσμα έναν χάρτη χαρακτηριστικών εξόδου 3×3 (9 εικονοστοιχεία).

Οι κύριες μορφές συναρτήσεων που χρησιμοποιούνται στο επίπεδο συγκέντρωσης είναι:

- **Max pooling:** Σε κάθε περιοχή του χάρτη χαρακτηριστικών, επιλέγεται η μέγιστη τιμή. Αυτό σημαίνει ότι μόνο η μέγιστη τιμή στην περιοχή του pooling διατηρείται, ενώ οι υπόλοιπες τιμές αγνοούνται.

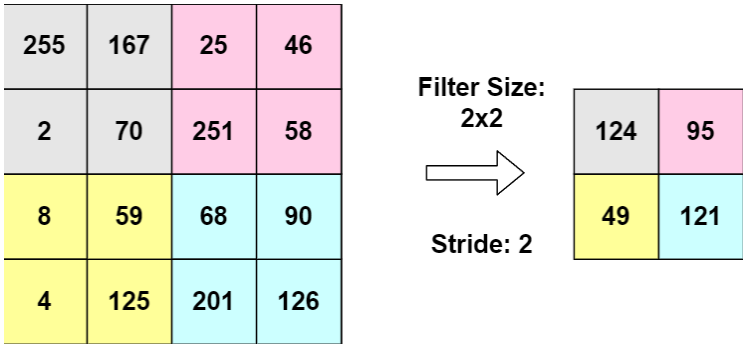
Εικόνα 11: Εφαρμογή max pooling με stride=2 και φίλτρα μεγέθους 2x2

- **Average pooling:** Σε κάθε περιοχή του χάρτη χαρακτηριστικών, υπολογίζεται ο μέσος όρος των τιμών και διατηρείται για αυτήν την περιοχή.



Εικόνα 12: Εφαρμογή average pooling με stride=2 και φίλτρα μεγέθους 2x2

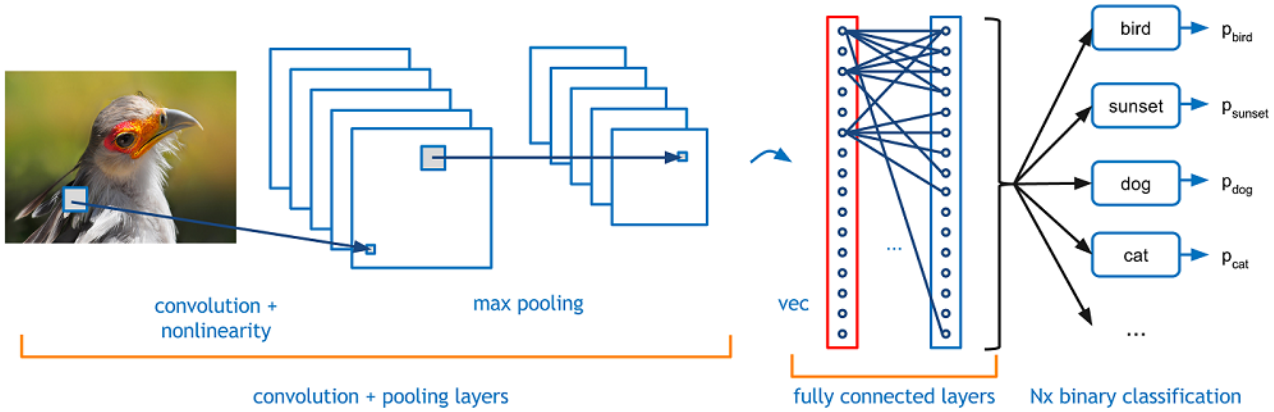
- **Min pooling:** Σε κάθε περιοχή του χάρτη χαρακτηριστικών, επιλέγεται η ελάχιστη τιμή. Αυτό σημαίνει ότι μόνο η ελάχιστη τιμή στην περιοχή του pooling διατηρείται. Χρησιμοποιείται κυρίως όταν η εικόνα έχει ανοιχτό χρώμα background, καθώς η ελάχιστη συγκέντρωση θα επιλέξει πιο σκούρα pixel.



Εικόνα 13: Εφαρμογή min pooling με stride=2 και φίλτρα μεγέθους 2x2

2.4.4 Fully Connected layer

Το **fully connected layer**(πλήρως συνδεδεμένο επίπεδο), FC, σε ένα συνελικτικό νευρωνικό δίκτυο(CNN) είναι ένα επίπεδο όπου κάθε νευρώνας συνδέεται με κάθε νευρώνα στο προηγούμενο επίπεδο. Το πλήρως συνδεδεμένο επίπεδο τοποθετείται στο τελευταίο επίπεδο του CNN και χρησιμοποιείται για την επιπεδοποίηση(flatten) των αποτελεσμάτων πριν την ταξινόμηση. Ουσιαστικά, ο βασικός στόχος του επιπέδου FC είναι να συνδυάσει τα χαρακτηριστικά που έχει μάθει προηγουμένως το δίκτυο προκειμένου να προβλέψει με μεγαλύτερη ακρίβεια την κλάση ταξινόμησης.



Εικόνα 14: Παράδειγμα κατηγοριοποίησης εικόνας. Στο τέλος φαίνονται τα Fully Connected Layers.

2.5 Υπερπαραμέτροι(hyperparameters) των στρωμάτων των CNN

Τα CNN διαφοροποιούνται μεταξύ τους ανάλογα με την αρχιτεκτονική τους, δηλαδή με τον τρόπο με τον οποίο τοποθετούνται τα επίπεδα σε σειρά, αλλά και με τις υπερπαραμέτρους που χρησιμοποιούνται στα Convolutional και Pooling layers. Οι παράμετροι που αναφέρθηκαν παραπάνω για τα επίπεδα της συνέλιξης και της συγκέντρωσης καθορίζονται πριν την έναρξη της εκπαίδευσης του δικτύου. Το μέγεθος των εξόδων των χαρτών χαρακτηριστικών εξαρτάται από αυτές τις υπερπαραμέτρους.

Συγκεντρωτικά αυτές είναι:

Για το Convolutional layer:

1. Ο αριθμός των φίλτρων που χρησιμοποιούνται.
2. Το μέγεθος των φίλτρων.
3. Το βήμα(stride).
4. Το zero-padding.

Για το Pooling layer:

1. Το μέγεθος των pooling κελιών.
2. Το βήμα μεταξύ των κελιών.

Ωστόσο, εξίσου σημαντικές είναι και οι υπερπαραμέτροι που αφορούν την διαδικασία εκπαίδευσης, όπως:

- Ο ρυθμός μάθησης.
- Ο αριθμός των εποχών εκπαίδευσης.
- Οι συναρτήσεις ενεργοποίησης που θα χρησιμοποιηθούν.
- Το batch size, δηλαδή ο αριθμός των υποδεδομένων που θα δίνονται στο δίκτυο κάθε φορά πριν την πραγματοποίηση της ενημέρωσης των παραμέτρων.

Η ρύθμιση όλων των υπερπαραμέτρων είναι ένα από τα σημαντικότερα ζητήματα στο χώρο των Συνελικτικών Νευρωνικών Δικτύων και παίζει καθοριστικό ρόλο στην λειτουργία τους.

Κεφάλαιο 3. Κατάτμηση εικόνας (Image Segmentation)

3.1 Ορισμός

Στην ψηφιακή επεξεργασία εικόνας και την υπολογιστική όραση, η κατάτμηση εικόνας είναι η διαδικασία τμηματοποίησης μιας ψηφιακής εικόνας σε πολλαπλές περιοχές. Ο στόχος της κατάτμησης είναι να απλοποιήσει ή/και να αλλάξει την αναπαράσταση μιας εικόνας σε κάτι πιο ουσιαστικό και πιο εύκολο στην ανάλυση. Η κατάτμηση εικόνας χρησιμοποιείται συνήθως για τον εντοπισμό αντικειμένων και ορίων (γραμμές, καμπύλες κ.λπ.) στις εικόνες. Πιο συγκεκριμένα, η κατάτμηση είναι η διαδικασία της ανάθεσης μιας ετικέτας σε κάθε εικονοστοιχείο έτσι ώστε εικονοστοιχεία με την ίδια ετικέτα να μοιράζονται ορισμένα κοινά χαρακτηριστικά.

Το αποτέλεσμα της κατάτμησης εικόνας είναι ένα σύνολο τμημάτων που ονομάζονται **superpixels** και καλύπτουν συλλογικά ολόκληρη την εικόνα. Κάθε ένα από τα εικονοστοιχεία σε ένα superpixel είναι παρόμοια σε σχέση με κάποια χαρακτηριστική ή υπολογίσιμη ιδιότητα, όπως το χρώμα, η ένταση ή η υφή.

3.2 Εφαρμογές της κατάτμησης εικόνας

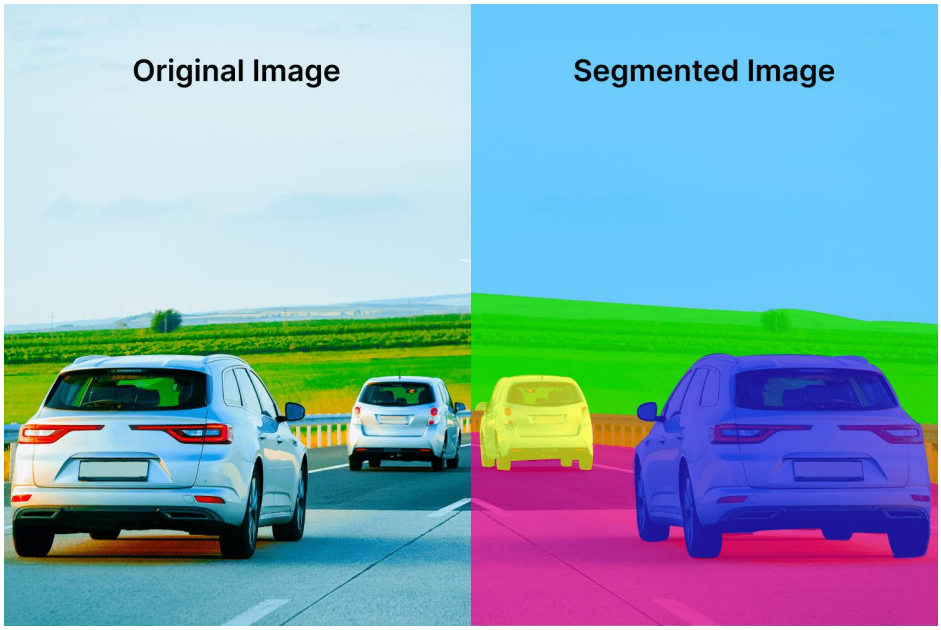
Η κατάτμηση εικόνας είναι μια ιδιαίτερα σημαντική διαδικασία στην υπολογιστική όραση με πολλές εφαρμογές σε διάφορους τομείς της ζωής και της επιστήμης. Η κατάτμηση εικόνας παίζει κρίσιμο ρόλο σε εφαρμογές ιατρικής απεικόνισης, όπως η ανίχνευση όγκων, η κατάτμηση οργάνων και η διάγνωση ασθενειών. Η ακριβής κατάτμηση των ανατομικών δομών και ανωμαλιών βοηθά στον χειρουργικό σχεδιασμό, την αξιολόγηση της θεραπείας και τη διάγνωση με τη βοήθεια του υπολογιστή.

Επιπλέον, τα τελευταία χρόνια υπάρχει μεγάλη ανάπτυξη στα συστήματα αυτόνομης οδήγησης. Σε αυτό τον τομέα η κατάτμηση εικόνας χρησιμοποιείται για την ανίχνευση

εμποδίων, κατακερματίζοντας αντικείμενα και το περιβάλλον, βοηθώντας τα αυτόνομα συστήματα στο να λαμβάνουν τεκμηριωμένες αποφάσεις και να πλοηγούνται με ασφάλεια.

Άλλες σημαντικές εφαρμογές είναι οι εξής:

- αναγνώριση προσώπου
- αναγνώριση αριθμού πινακίδων
- ανάλυση δορυφορικών εικόνων
- αυτό-οδηγούμενα οχήματα
- αυτόματη επεξεργασία εικόνας και βίντεο



Εικόνα 15: Παράδειγμα κατάτμησης εικόνας σε εφαρμογή αυτόνομης οδήγησης. Η δεξιά εικόνα παρουσιάζει το αποτέλεσμα της κατάτμησης.

3.3 Παραδοσιακές τεχνικές κατάτμησης

Για την κατάτμηση εικόνων χρησιμοποιούνται διάφορες "παραδοσιακές" τεχνικές, όπως η κατάτμηση με βάση την ανίχνευση ακμών, η κατάτμηση με μεθόδους βασισμένες σε ιστογράμματα, η κατάτμηση με χρήση τεχνικών κατωφλίωσης, αλλά και χρησιμοποιώντας μεθόδους ομαδοποίησης π.χ. k-means. Τέτοιες παραδοσιακές μέθοδοι αποδεικνύονται λιγότερο αποτελεσματικές σε σύγκριση με τις πιο προηγμένες τεχνικές βαθιάς μάθησης. Αυτό οφείλεται στο γεγονός ότι χρησιμοποιούν αλγόριθμους με μικρό βαθμό ευελιξίας και απαιτούν συνήθως την παρέμβαση ενός ανθρώπου με ειδίκευση, ώστε να λειτουργήσουν ορθά.

Στην παρούσα εργασία υλοποιείται ένας αλγόριθμος που συνδυάζει χαρακτηριστικά της βαθιάς μάθησης με τεχνικές ομαδοποίησης.

3.3.1 Μέθοδος ομαδοποίησης k-means

Ομαδοποίηση: Η ομαδοποίηση είναι μια διαδικασία μάθησης χωρίς επίβλεψη που στόχο έχει την διαίρεση του συνόλου δεδομένων σε ομάδες, έτσι ώστε δεδομένα στην ίδια ομάδα να είναι παρόμοια ή να σχετίζονται μεταξύ τους, καθώς και να διαφέρουν από τα δεδομένα των άλλων ομάδων. Σκοπός της ομαδοποίησης είναι να ανακαλύψει μοτίβα ή δομές στα δεδομένα χωρίς προηγούμενη γνώση των κατηγοριών στις οποίες ανήκουν τα δεδομένα.

Ο αλγόριθμος k-means [15] είναι ένας αλγόριθμος ομαδοποίησης (clustering) που χρησιμοποιείται για την ομαδοποίηση ενός συνόλου δεδομένων σε ομάδες (clusters) με βάση τις ομοιότητες μεταξύ των δεδομένων. Ο αλγόριθμος k-means χρησιμοποιείται και για τον διαχωρισμό των εικονοστοιχείων σε k ομάδες ακολουθώντας τα βήματα:

1. **Αρχικοποίηση:** Επιλέγουμε Κ κέντρα από το σύνολο των δεδομένων είτε τυχαία είτε χρησιμοποιώντας κάποια ευρετική μέθοδο όπως ο k-means++[9].
2. **Ανάθεση:** Σε κάθε pixel ανατίθεται ως κέντρο του, το κοντινότερο σε αυτό κέντρο, δηλαδή σε αυτό που ελαχιστοποιεί την απόσταση μεταξύ του pixel και κέντρου.
3. **Ενημέρωση των κέντρων:** Αφού σε όλα τα pixels έχει ανατεθεί κάποιο cluster, επαναυπολογίζονται ξανά τα κέντρα των clusters ως το μέσο των pixels του cluster.
4. **Επανάληψη:** Τα δύο προηγούμενα βήματα επαναλαμβάνονται μέχρι να συγκλίνει ο αλγόριθμος. Συνήθως, η σύγκλιση επιτυγχάνεται όταν τα κέντρα των clusters παραμένουν σταθερά μεταξύ των επαναλήψεων ή όταν ο αλγόριθμος φτάσει σε ένα προκαθορισμένο αριθμό επαναλήψεων.

Η απόσταση που υπολογίζεται μεταξύ των pixel και των κέντρων είναι η ευκλείδεια απόσταση και συνήθως υπολογίζεται με βάση το χρώμα είτε την φωτεινότητα είτε την θέση είτε συνδυασμούς αυτών. Η ποιότητα της λύσης του αλγορίθμου εξαρτάται σε μεγάλο βαθμό από την αρχικοποίηση τους και τον αριθμό των κέντρων. Συγκεκριμένα, μία τεχνική που μπορεί να χρησιμοποιηθεί για την αξιολόγηση της ποιότητας της ομαδοποίησης είναι το **silhouette score.**

3.3.1.1 Silhouette score

Το silhouette score [4] χρησιμοποιείται για την αξιολόγηση της ποιότητας των κέντρων που δημιουργούνται από αλγόριθμους ομαδοποίησης όπως ο k-means. Η μετρική αυτή υπολογίζει το πόσο κοντά βρίσκεται κάθε σημείο του cluster στα γειτονικά σημεία των άλλων clusters. Οι τιμές που υπολογίζει κυμαίνονται από -1 έως +1, όπου τιμές κοντά στο +1 υποδεικνύουν ότι το σημείο βρίσκεται μακριά από τα γειτονικά clusters και επομένως έχει ανατεθεί στο σωστό cluster. Τιμές μικρότερες του 0 υποδεικνύουν ότι είτε το σημείο βρίσκεται πολύ κοντά σε όριο απόφασης μεταξύ δύο γειτονικών κέντρων είτε έχει ανατεθεί σε λάθος cluster. Για να υπολογιστεί το silhouette score για κάθε σημείο, πρέπει να βρεθούν οι ακόλουθες αποστάσεις:

1. Η μέση απόσταση μεταξύ του σημείου και όλων των άλλων σημείων στο ίδιο cluster. Αυτή η απόσταση μπορεί επίσης να ονομαστεί ως μέση απόσταση εντός cluster. Η μέση αυτή απόσταση συμβολίζεται με α.
2. Η μέση απόσταση μεταξύ του σημείου και όλων των άλλων σημείων του πλησιέστερου cluster, του οποίου δεν είναι μέλος. Αυτή η απόσταση μπορεί επίσης να ονομαστεί ως μέση απόσταση σημείου-πλησιέστερου cluster. Η μέση απόσταση συμβολίζεται με b.

Το silhouette score υπολογίζεται με τον ακόλουθο τύπο:

$$(b - a) / \max(a, b)$$

Έτσι, υπολογίζοντας το silhouette score από όλα τα pixels και για διάφορα k μπορούμε να επιλέξουμε το βέλτιστο αριθμό ομάδων.

3.4 Σημασιολογική κατάτμηση εικόνας μέσω τεχνικών deep learning

Η σημασιολογική κατάτμηση εικόνας στοχεύει στην αντιστοίχιση σημασιολογικών ετικετών σε κάθε pixel μιας εικόνας, παρέχοντας μια λεπτομερή κατανόηση της σκηνής.

Διαφέρει από άλλες τεχνικές κατάτμησης εστιάζοντας στην επισήμανση σε επίπεδο αντικειμένου και όχι μεμονωμένων περιοχών pixels. Οι σύγχρονες τεχνικές κατάτμησης χρησιμοποιούν σε μεγάλο βαθμό βαθιά μάθηση(deep learning) και η χρήση των Συνελικτικών Νευρωνικών Δικτύων έχει αποδεχθεί ιδιαίτερα αποτελεσματική για το πρόβλημα αυτό.

Σε σύγκριση με τα προβλήματα ταξινόμησης (classification) και εντοπισμού (detection), η σημασιολογική κατάτμηση (semantic segmentation) είναι μια πολύ πιο δύσκολη εργασία. Επιπλέον, οι μέθοδοι σημασιολογικής κατάτμησης διαφέρουν από τις απλές τεχνικές κατάτμησης στο ότι χρησιμοποιούν κυρίως τεχνικές επιβλεπόμενης μάθησης.

Γενικά, η **Semantic Segmentation**: Ταξινομεί την κατηγορία για κάθε pixel μέσα σε μια εικόνα.

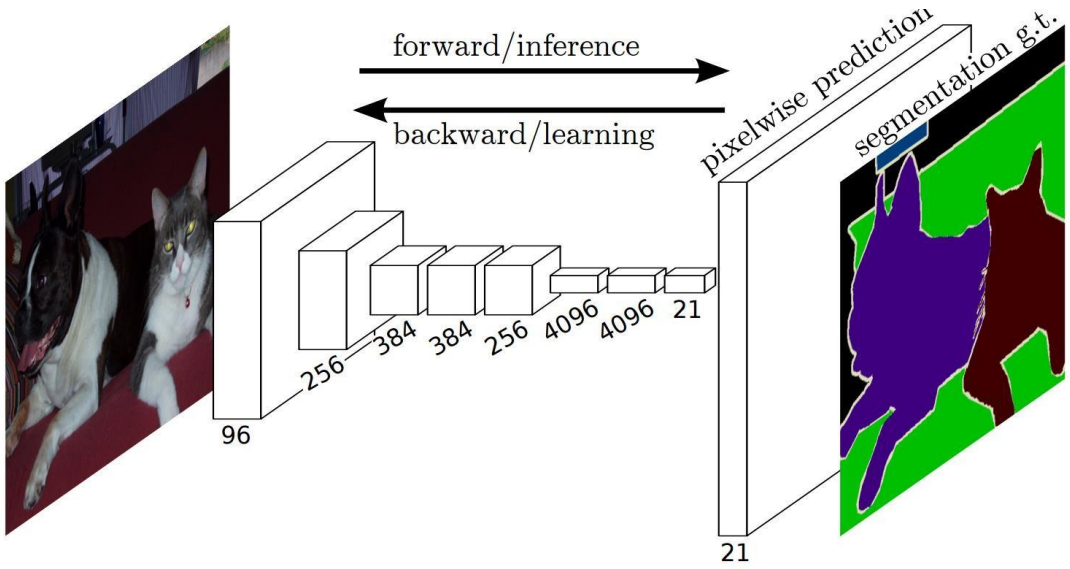
3.4.1 Fully Convolutional Networks for semantic segmentation

Τα **Fully Convolutional Networks** (FCN) είναι ένα είδος νευρωνικών δικτύων που έχουν σχεδιαστεί κυρίως για προβλήματα με εικόνες και την αντιμετώπιση προβλημάτων υπολογιστικής όρασης, όπως η σημασιολογική κατάτμηση και η αναγνώριση αντικειμένων. Η βασική ιδέα πίσω από τα FCNs είναι να αντικαταστήσουν τα πλήρως συνδεδεμένα επίπεδα των συμβατικών CNNs που δεν μπορούν να διαχειριστούν διαφορετικά μεγέθη εισόδου, με συνελικτικά επίπεδα (convolutional layers) που μπορούν να χειριστούν διαφορετικές διαστάσεις εισόδου. Έτσι, η διαδικασία εκπαίδευσης και πρόβλεψης μπορεί να εφαρμοστεί απευθείας σε εικόνες οποιουδήποτε μεγέθους, σε γρήγορο ρυθμό και χωρίς την ανάγκη για προεπεξεργασία ή αλλαγή μεγέθους. Το τελευταίο επίπεδο εξόδου έχει ένα μεγάλο δεκτικό πεδίο(receptive field), που αντιστοιχεί στο ύψος και το πλάτος της εικόνας, ενώ ο αριθμός των καναλιών αντιστοιχεί στον αριθμό των κατηγοριών. Στο επίπεδο συνέλιξης κάθε εικονοστοιχείο ταξινομείται ξεχωριστά.

Εικόνα 16: Παράδειγμα εικόνας που έχει προκύψει από FCN.

Στην 1η εικόνα είναι η αρχική εικόνα εισόδου, στην 2η εικόνα είναι η ground truth εικόνα, στην 3η η έξοδος του δικτύου με την κατάτμηση των αντικειμένων, στην 4η η επικάλυψη της αρχικής εικόνας με την έξοδο.

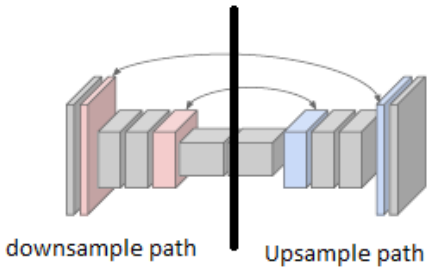
Οι εργασίες [2] και [3] παρουσιάζουν την δομή και την λειτουργία των FCN.
Κατά την ταξινόμηση(classification), η εικόνα εισόδου συρρικνώνεται καθώς περνάει από τα convolutional και pooling επίπεδα εξάγοντας χαρακτηριστικά της και τελικά σαν αποτέλεσμα παίρνουμε τις ετικέτες που αντιπροσωπεύουν την εικόνα εισόδου. Αν υποθέσουμε ότι αντικαθιστούμε τα FC layers με convolutional layers και η αρχική εικόνα δεν συρρικνωθεί τότε το αποτέλεσμα δεν θα είναι μια απλή ετικέτα, αλλά θα έχει κάποιο μέγεθος, μικρότερο βέβαια της αρχικής εικόνας. Στη συνέχεια αν μεγεθυνθεί το αποτέλεσμα, μπορεί να υπολογιστεί ένα pixelwise prediction, η κατατμημένη εικόνα.



Εικόνα 17: Παράδειγμα FCN για σημασιολογική κατάτμηση.

Η αρχιτεκτονική των FCN μπορεί να χωριστεί σε δύο μέρη:

1. Το Downsampling μέρος, δηλαδή το μέρος όπου τα επίπεδα που το απαρτίζουν μειώνουν την διάσταση της εισόδου, χάνοντας έτσι χωρική πληροφορία με αντάλλαγμα όμως την εύρεση χρήσιμων χαρακτηριστικών για την αποτελεσματική διάκριση των διαφορετικών αντικειμένων.
2. Το Upsampling μέρος, όπου γίνεται η ανάκτηση της χωρικής πληροφορίας που χάθηκε.

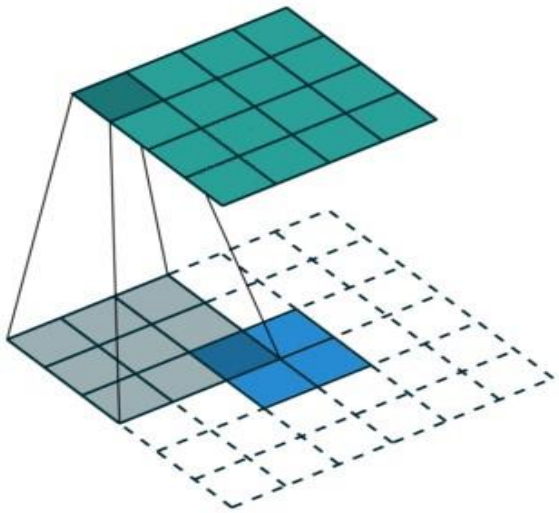


Εικόνα 18: Στην αριστερή πλευρά φαίνεται το downsampling μέρος, όπου κάθε κουτάκι εκφράζει το μέγεθος της διάστασης εξόδου του επιπέδου, ενώ δεξιά φαίνεται το upsampling μέρος.

Άλλη μία τεχνική που χρησιμοποιείται για την αντιμετώπιση του προβλήματος της απώλειας χωρικής πληροφορίας από τα downsample επίπεδα είναι η χρήση **skip connections**. Συγκεκριμένα, τα skip connections επιτρέπουν την παράκαμψη (skip) συνελικτικών επιπέδων. Στα FCN χρησιμοποιείται συχνά για τη μεταφορά τοπικών πληροφοριών συνδυάζοντας ή αθροίζοντας χάρτες χαρακτηριστικών από το Downsampling path με χάρτες χαρακτηριστικών από το Upsampling path. Αυτές οι συνδέσεις παράβλεψης(skip connections) παρέχουν αρκετές πληροφορίες σε μεταγενέστερα επίπεδα για τη δημιουργία των ορίων κατάτμησης.

3.4.1.1 Upsampling via deconvolution(Upsampling μέσω αποσυνέλιξης)

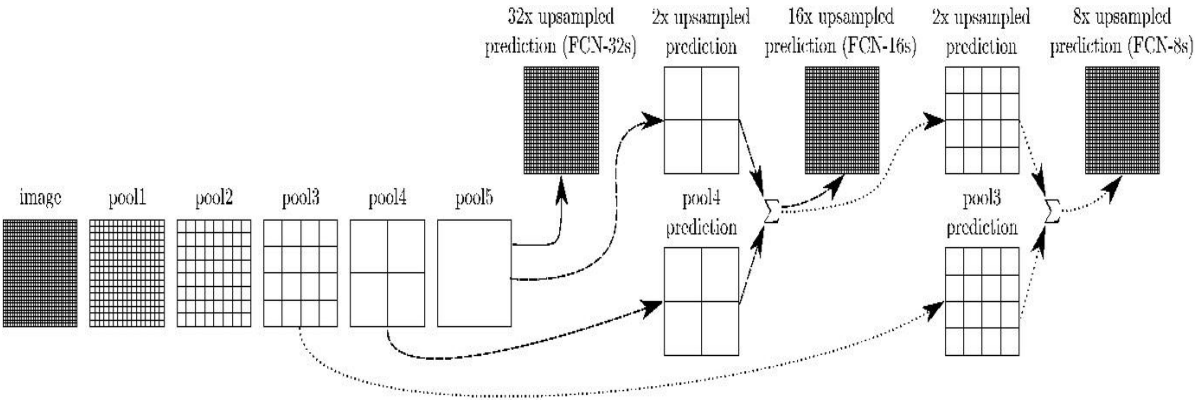
Το όνομα αποσυνέλιξη(deconvolution) που χρησιμοποιείται στο upsampling path, προέρχεται από το γεγονός ότι κάνουμε την αντίθετη διαδικασία από το επίπεδο της συνέλιξης, δηλαδή αυξάνουμε το μέγεθος της εξόδου. Ωστόσο, η αποσυνέλιξη δεν θα πρέπει να παρερμηνεύεται ως το ακριβώς αντίστροφο της συνέλιξης, πράγμα που δεν είναι. Επίσης, η διαδικασία αυτή είναι ακόμα γνωστή και ως up convolution, transposed convolution και fractional stride convolution όταν χρησιμοποιείται κλασματικό βήμα(fractional stride).



Εικόνα 19: Upsampling via Deconvolution(η μπλε περιοχή είναι η είσοδος, η γκρι περιοχή το φίλτρο αποσυνέλιξης και η πράσινη η έξοδος)

3.4.1.2 Συγχώνευση των εξόδων("Fusing the output")

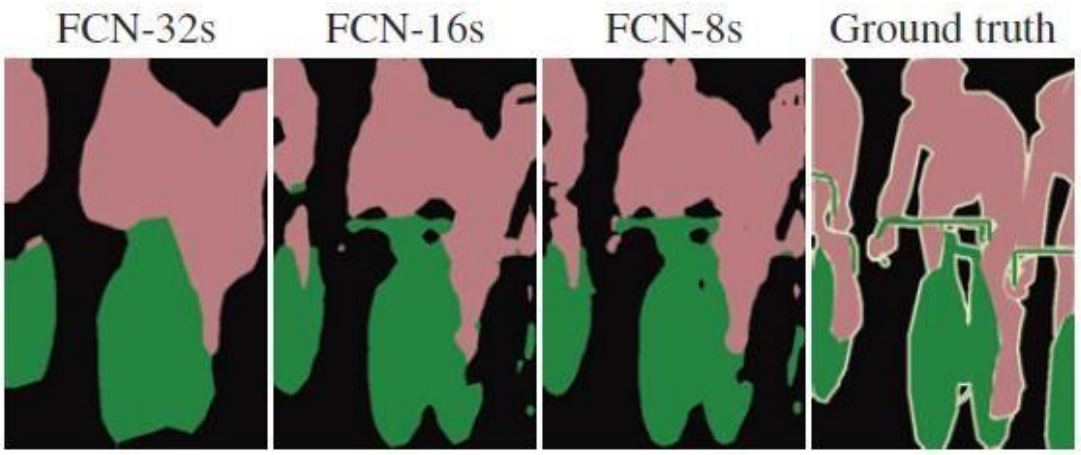
Η διαδικασία της συγχώνευσης των εξόδων("**fusing the output**") στα FCN αναφέρεται στο συνδυασμό των εξόδων που προκύπτουν από διαφορετικά επίπεδα του μοντέλου με σκοπό την αύξηση των χωρικών πληροφοριών στα βαθύτερα επίπεδα. Έτσι, εάν συνδυάσουμε, συγχωνεύσουμε τα χαρακτηριστικά εξόδου ενός βαθύτερου επιπέδου με τα χαρακτηριστικά εξόδου από τα αρχικά επίπεδα, μπορούμε να βελτιώσουμε το αποτέλεσμα. Υπάρχουν παραλλαγές της αρχιτεκτονικής FCN, οι οποίες διαφέρουν κυρίως στην χωρική ακρίβεια της εξόδου τους. Για παράδειγμα, οι παρακάτω εικόνες δείχνουν τις παραλλαγές FCN-32, FCN-16 και FCN-8.



Εικόνα 20: Παράδειγμα εφαρμογής της συγχώνευσης εξόδων στο FCN-8 και FCN-16

- FCN-32: Παράγει άμεσα το segmentation map από το conv7, χρησιμοποιώντας ένα deconvolution επίπεδο με stride 32.
- FCN-16: Αθροίζει την πρόβλεψη 2× του δείγματος από το pool5 με το pool4 και στη συνέχεια παράγει το segmentation map, χρησιμοποιώντας ένα deconvolution επίπεδο με stride 16 πάνω από αυτό.
- FCN-8: Παρόμοια με το FCN-16 όπως φαίνεται στην εικόνα 20.

Στην παρακάτω εικόνα φαίνεται και η ποιοτική διαφορά των τριών διαφορετικών αρχιτεκτονικών FCN, παίρνοντας το αποτέλεσμα που βγάζουν για την ίδια εικόνα. Βλέπουμε ότι στο FCN-32 το αποτέλεσμα είναι πιο χοντροκομμένο και λιγότερο αυστηρό λόγω της απώλειας χωρικής πληροφορίας, ενώ στα υπόλοιπα έχουμε καλύτερα αποτελέσματα.

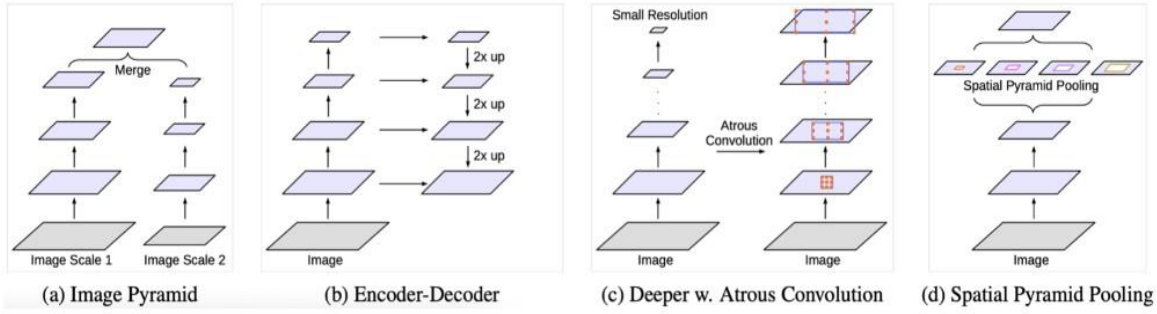


Εικόνα 21: Σύγκριση αποτελεσμάτων μεταξύ διαφορετικών FCNs.

3.4.2 DeepLabV3

Το DeepLabV3 [8] είναι μια προηγμένη αρχιτεκτονική Βαθιών Συνελικτικών Νευρωνικών Δικτύων(DCNN) που αναπτύχθηκε από την Google για την επίλυση προβλημάτων σημασιολογικής κατάτμησης εικόνας και αναγνώρισης προτύπων. Στην εργασία [8] επισημαίνονται οι διάφοροι αλγόριθμοι που έχουν προταθεί σε σχετικές εργασίες για την αναγνώριση αντικειμένων ανεξαρτήτως της κλίμακας τους όπως:

1. Image pyramid
2. Encoder-Decoder
3. Deeper w. Atrous Convolution
4. Spatial Pyramid Pooling



Εικόνα 22: Alternative architectures to capture multi-scale context.

Το μοντέλο που προτείνεται στο DeepLav3 συνδυάζει τα 3. και 4. και θα αναλυθεί στην συνέχεια.

Το DeepLabv3 χρησιμοποιεί ένα προ-εκπαιδευμένο ResNet ως το κύριο δίκτυο εξαγωγής χαρακτηριστικών. Εκτός από αυτό, ωστόσο, προτείνεται ένα νέο μπλοκ Residual για την εκμάθηση χαρακτηριστικών πολλαπλών κλιμάκων. Αντί της κλασικής συνέλιξης, το τελευταίο μπλοκ Residual χρησιμοποιεί συνέλιξη **Atrous(atrous convolutions)**, όπου κάθε συνέλιξη χρησιμοποιεί διαφορετικό dilation rate για να καταγράψει πληροφορίες πολλαπλών κλιμάκων. Τέλος, το DeepLabv3 χρησιμοποιεί Atrous **Spatial Pyramid Pooling** (ASPP) πάνω από αυτό το νέο μπλοκ. Η ASPP χρησιμοποιεί Atrous

convolutions με διαφορετικά dilation rate για την ταξινόμηση περιοχών αυθαίρετης κλίμακας. Έτσι, τα τρία βασικά συστατικά που συνθέτουν την αρχιτεκτονική του DeepLab είναι:

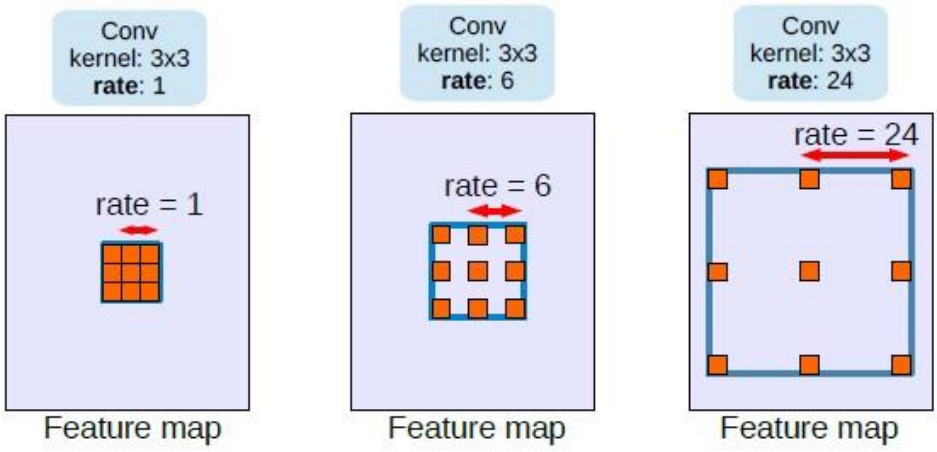
1. Η αρχιτεκτονική του ResNet
2. Η Atrous convolution
3. Η Atrous Spatial Pyramid Pooling (ASPP)

3.4.2.1 Atrous Convolution

Ο επαναλαμβανόμενος συνδυασμός φίλτρων συνέλιξης με pooling layers σε διαδοχικά επίπεδα, όπως είδαμε, μειώνει σημαντικά τη χωρική πληροφορία στις εξόδους των χαρτών χαρακτηριστικών που προκύπτουν. Έτσι, στο DeepLab εισάγονται **Dilated** ή **Atrous Convolutions**. Οι Atrous convolutions είναι σαν τις κλασικές συνελίξεις, με τη διαφορά ότι εισάγονται κενά(atrous) μεταξύ των στοιχείων των φίλτρων, με αποτέλεσμα την επέκταση του οπτικού πεδίου του φίλτρου και την περαιτέρω λήψη χωρικής πληροφορίας. Η παράμετρος που καθορίζει το μήκος του κενού ονομάζεται **dilation rate**, συμβολίζεται με r και καθορίζει την ποσότητα χωρικής πληροφορίας που το φίλτρο μπορεί να "πιάσει" μέσω του οπτικού του πεδίου. Ο τύπος της atrous convolution δίνεται ως εξής:

$$y[i] = \sum_k x[i + r \times k]w[k]$$

Η κλασική συνέλιξη χρησιμοποιεί dilation rate r=1. Αν θέσουμε το r σε μεγαλύτερη τιμή αυτό θα έχει ως αποτέλεσμα την διεύρυνση του φίλτρου συνέλιξης και του οπτικού πεδίου(receptive field) του. Έτσι, με την τιμή του r ουσιαστικά εισάγουμε r-1 μηδενικά ανάμεσα σε δύο τιμές του φίλτρου.



Εικόνα 23: Atrous Convolutions με διαφορετικά dilation rates.

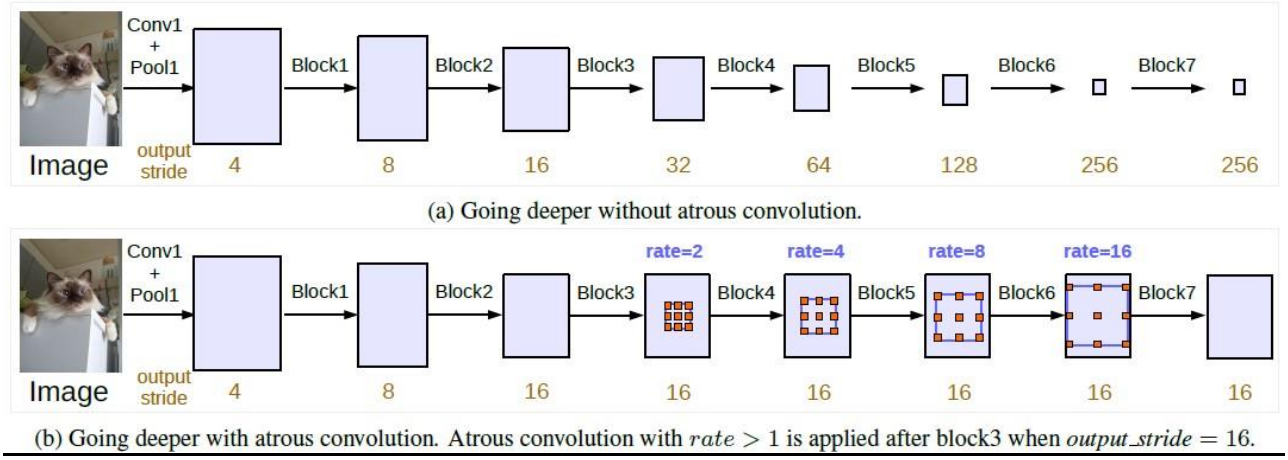
Τα πλεονεκτήματα της Atrous Convolution είναι:

- Επιτρέπει τον έλεγχο του οπτικού πεδίου(receptive field) του φίλτρου.
- Διατηρεί χωρική πληροφορία.
- Ελέγχει την πυκνότητα της εξαγωγής χαρακτηριστικών στα βαθύτερα layers.
- Διατηρεί τον ίδιο αριθμό υπερπαραμέτρων με την κλασική συνέλιξη.

3.4.2.2 Going deeper with Atrous Convolution

Όπως είδαμε η επιλογή του κατάλληλου dilation rate r επηρεάζει σε μεγάλο βαθμό την αποτελεσματικότητα του δικτύου.

Επομένως, είναι σημαντικό να γνωρίζουμε την αναλογία του μεγέθους της εικόνας εισόδου προς το μέγεθος της εξόδου του χάρτη χαρακτηριστικών(output feature map). Η αναλογία αυτή είναι γνωστή με το όνομα **output stride.** Για μια εικόνα εισόδου διάστασης 224x224x3, όπου το ύψος και πλάτος ισούνται με 224pixels και χρησιμοποιώντας τα 3 κανάλια rgb, τότε αν έχουμε μέγεθος του χάρτη εξόδου ίσο με 14x14 θα έχουμε output stride=224/14=16. Στην δημοσίευση του DeepLab τονίζεται ότι τα μοντέλα με μικρότερο output stride τείνουν να παράγουν καλύτερα αποτελέσματα κατάτμησης, ωστόσο απαιτούν περισσότερο χρόνο για την εκπαίδευση τους.



Εικόνα 24: Στην πάνω εικόνα (a) παρουσιάζεται ένα μοντέλο χωρίς χρήση atrous convolutions, ενώ στην κάτω εικόνα (b) γίνεται χρήση αυτών.

Στην εικόνα (a) πραγματοποιούνται κλασικές συνελίξεις(r=1) και συγκεντρώσεις(pooling) έχοντας ως αποτέλεσμα την αύξηση του output stride όσο προχωράμε στα βαθύτερα επίπεδα. Έτσι, χωρικές πληροφορίες χάνονται, με συνέπεια την μείωση της ποιότητας της σημασιολογικής κατάτμησης.

Στην εικόνα (b) που πραγματοποιούνται atrous convolutions με r>1 μετά το block 3, μπορούμε να διακρίνουμε ότι αυξάνοντας το οπτικό πεδίο, δηλαδή διπλασιάζοντας το r σε κάθε επόμενο επίπεδο καταφέρνουμε να διατηρούμε το μέγεθος της εξόδου του χάρτη χαρακτηριστικών σταθερό.

Το δίκτυο DeepLab εισάγει την έννοια του πολλαπλού πλέγματος(multi grid), όπου διαφορετικά dilation rates εφαρμόζονται σε διαφορετικά blocks του δικτύου, όπως φαίνεται από το block 4 έως 7.

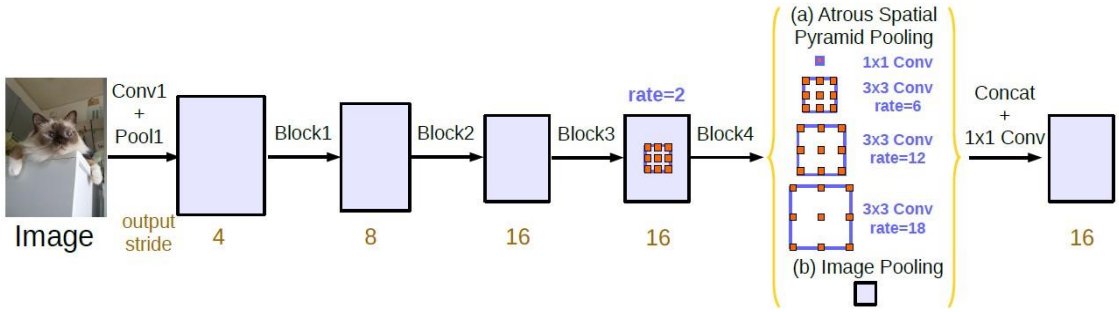
Στο παραπάνω παράδειγμα , όταν θέλουμε output stride = 16 και έχουμε multi grid=(1, 2, 4), οι τρεις συνελίξεις θα έχουν dilation rates = 2x(1, 2, 4)=(2, 4, 8) στα block 4 έως 6.

3.4.2.3 Atrous Spatial Pyramid Pooling(ASPP)

Το Atrous Spatial Pyramid Pooling (ASPP) είναι μια τεχνική εξαγωγής χαρακτηριστικών που εισήχθη για πρώτη φορά στο δίκτυο DeepLab. Το ASPP εφαρμόζει ένα σύνολο πα-

ράλληλων διεσταλμένων συνελίξεων(dilated/atrous convolutions) με διαφορετικά dilation rates για την εξαγωγή χαρακτηριστικών σε **διαφορετικές κλίμακες**. Στη συνέχεια, η έξοδος κάθε atrous convolution συγκεντρώνεται μέσω καθολικών λειτουργιών συγκέντρωσης(pooling), οι οποίες βοηθούν στη λήψη πληροφοριών σε διαφορετικές κλίμακες. Οι χάρτες χαρακτηριστικών εξόδου κάθε παράλληλης διαδρομής συνενώνονται και τροφοδοτούνται για επεξεργασία σε ένα συνελικτικό επίπεδο 1x1 με 256 φίλτρα και batch normalization. Έπειτα, η έξοδος από αυτό το επίπεδο περνάει από μία τελική συνέλιξη που παράγει τα τελικά χαρακτηριστικά.

Η διαδικασία αυτή επιτρέπει στο δίκτυο να συλλαμβάνει τόσο τοπικές όσο και global πληροφορίες της εικόνας, οι οποίες μπορεί να είναι κρίσιμες για την ακριβή κατάτμηση εικόνων με δεδομένα σε πολλαπλές κλίμακες. Ο λόγος χρήσης του ASPP είναι ότι ανακαλύφθηκε ότι καθώς το dilation rate γινόταν μεγαλύτερο, ο αριθμός των έγκυρων βαρών του φίλτρου (δηλαδή, τα βάρη του φίλτρου που εφαρμόζονται στην έγκυρη περιοχή χαρακτηριστικών, αντί για τα μηδενικά του padding) γινόταν μικρότερος.



Εικόνα 25: Παράδειγμα χρήσης Atrous Spatial Pyramid Pooling(ASPP) layer μετά το block 4.

Στην εικόνα 25 στο επίπεδο ASPP περιέχονται 4 παράλληλες λειτουργίες atrous convolution με ένα φίλτρο 1x1 και τρία φίλτρα 3x3 με dilation rates r = (6, 12, 18) αντίστοιχα. Έπειτα, τα features που προκύπτουν τροφοδοτούνται σε μια συνέλιξη 1×1 με 256 φίλτρα όπου προκύπτουν τα τελικά χαρακτηριστικά της κατάτμησης..

Κεφάλαιο 4. Αξιοποίηση features από pre-trained CNN για κατάτμηση εικόνων

4.1 PyTorch και προ-εκπαιδευμένα μοντέλα για σημασιολογική κατάτμηση

Το PyTorch είναι μια ανοιχτού κώδικα βιβλιοθήκη μηχανικής μάθησης για Python που αναπτύχθηκε από το Facebook's AI Research lab (FAIR). Έχει κερδίσει μεγάλη δημοτικότητα λόγω της ευελιξίας, της ευκολίας χρήσης και της ισχυρής υποστήριξης που προσφέρει για την ανάπτυξη και την εκπαίδευση νευρωνικών δικτύων. Η βιβλιοθήκη χρησιμοποιείται για εφαρμογές κυρίως υπολογιστικής όρασης και επεξεργασίας φυσικής γλώσσας.
Η PyTorch προσφέρει τέσσερα προ-εκπαιδευμένα CNN για σημασιολογική κατάτμηση εικόνας. Τα μοντέλα αυτά είναι όλα εκπαιδευμένα πάνω στο γνωστό COCO train2017 dataset[6] και στις 20 κατηγορίες που παρουσιάζονται στο Pascal VOC dataset[5]. Τα μοντέλα είναι τα εξής:

1. FCN με backbone ResNet50
2. FCN με backbone ResNet101
3. DeepLabV3 με backbone ResNet50
4. DeepLabV3 με backbone ResNet101

Οι ακρίβειες αυτών των προ-εκπαιδευμένων μοντέλων που αξιολογήθηκαν στο dataset COCO val2017 [6] έχουν ως εξής:

Network	**Mean IoU**	**Global pixelwise acc**
FCN ResNet50	60.5	91.4
FCN ResNet101	63.7	91.9
DeepLabV3 ResNet50	66.4	92.4
DeepLabV3 ResNet101	67.4	92.4

Από τον παραπάνω πίνακα φαίνεται ότι το DeepLabV3 δίνει τα καλύτερα αποτελέσματα από τα υπόλοιπα.

Το Intersection over Union (IoU) είναι μια μετρική που προσδιορίζει ποσοτικά την επικάλυψη μεταξύ της εξόδου πρόβλεψης του κάθε μοντέλου με τον στόχο του ground truth. Το IoU παρέχει ένα μέτρο του πόσο καλά ευθυγραμμίζεται ένα αντικείμενο που προβλέφθηκε από το μοντέλο με την πραγματική απεικόνιση του αντικειμένου, επιτρέποντας την αξιολόγηση της ακρίβειας του μοντέλου και τη λεπτομερή ρύθμιση των αλγορίθμων για βελτιωμένα αποτελέσματα. Η μετρική αυτή υπολογίζεται διαιρώντας τον αριθμό των κοινών pixels μεταξύ των δύο μασκών με τον συνολικό αριθμό pixels που υπάρχουν και στις δύο μάσκες.

Το global pixelwise accuracy είναι μια μετρική που υπολογίζει το ποσοστό των εικονοστοιχείων που ταξινομούνται σωστά σε σχέση με το συνολικό αριθμό των εικονοστοιχείων.

4.2 Μελέτη του pre-trained DeepLabV3

Αρχικά, με τη βοήθεια της PyTorch χρησιμοποιήσαμε το προ-εκπαιδευμένο μοντέλο DeepLabV3 προκειμένου να καταλάβουμε την πλήρη λειτουργία του και τον τρόπο με τον οποίο επεξεργάζεται τις εικόνες που δέχεται σαν είσοδο. Κυρίως χρησιμοποιήσαμε ως backbone του DeepLabV3 το ResNet 101 καθώς αυτό το δίκτυο είναι ένα από τα state-of-the-art δίκτυα στο χώρο του semantic segmentation.

Οι κλάσεις που αναγνωρίζει το προ-εκπαιδευμένο μοντέλο είναι οι ακόλουθες:

1. ‘airplane’
2. ‘bicycle’
3. ‘bird’
4. ‘boat’
5. ‘bottle’
6. ‘bus’
7. ‘car’
8. ‘cat’
9. ‘chair’
10. ‘cow’
11. ‘dining table’
12. ‘dog’
13. ‘horse’
14. ‘motorbike’
15. ‘person’
16. ‘potted plant’
17. ‘sheep’
18. ‘sofa’
19. ‘train’
20. ‘tv monitor’

Ουσιαστικά, το δίκτυο DeepLabV3 ταξινομεί κάθε εικονοστοιχείο σε μία από τις 20 κατηγορίες, ενώ τα εικονοστοιχεία που δεν ανήκουν σε κάποια από αυτές θεωρούνται ως background και παίρνουν μια αυθαίρετη τιμή π.χ. 0.

Παρακάτω παρουσιάζονται τα αποτελέσματα της κατάτμησης από το δίκτυο DeepLabv3 χρησιμοποιώντας ως backbone το προεκπαιδευμένο ResNet101:

Εικόνα 26: Από αριστερά, 1η εικόνα Αρχική εικόνα εισόδου, 2η εικόνα έξοδος του DeepLabv3, 3η εικόνα επικάλυψη(overlay) του αντικειμένου που αναγνωρίστηκε με την αρχική εικόνα.

Εικόνα 27: Από αριστερά, 1η εικόνα Αρχική εικόνα εισόδου, 2η εικόνα έξοδος του DeepLabv3, 3η εικόνα επικάλυψη(overlay) του αντικειμένου που αναγνωρίστηκε με την αρχική εικόνα.

Στις εικόνες 26, 27 βλέπουμε ότι το δίκτυο εντοπίζει τα pixels που αντιστοιχούν στις γνωστές κατηγορίες σκύλος και άνθρωπος αντίστοιχα, ενώ τα υπόλοιπα pixels που δεν ανήκουν σε κάποια από τις 20 κατηγορίες ταξινομούνται ως background(μαύρο χρώμα).

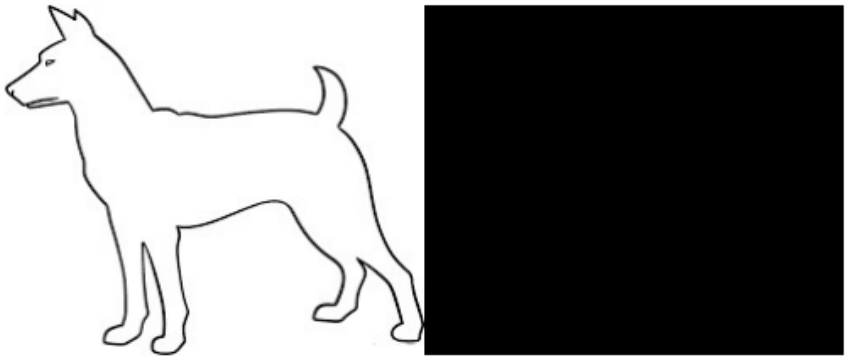
Εικόνα 28: Από αριστερά, (1η εικόνα) Αρχική εικόνα εισόδου, (2η εικόνα) έξοδος του DeepLabv3 όπου αναγνωρίστηκαν δύο διαφορετικές κατηγορίες αντικειμένων, (3η εικόνα) επικάλυψη(overlay) των αντικειμένων που αναγνωρίστηκαν με την αρχική εικόνα.

Στην εικόνα 28, η αρχική εικόνα 1 βλέπουμε ότι περιέχει αντικείμενα από δύο γνωστές κατηγορίες (person και sheep) και αυτό αποτυπώνεται με επιτυχία στο αποτέλεσμα καθώς οι διαφορετικές κατηγορίες αποτυπώθηκαν με διαφορετικά χρώματα. Επιπλέον, το DeepLabV3 κατάφερε να εντοπίσει και τα τέσσερα πρόβατα της εικόνας, αλλά απέτυχε να αναγνωρίσει τον σκύλο ανάμεσα τους και ταξινόμησε τα pixels του ως sheep. Όπως και πριν όλα τα υπόλοιπα «άγνωστα» pixels ταξινομούνται από το δίκτυο ως background.

4.2.1 Χωρική ομαλότητα DeepLabV3

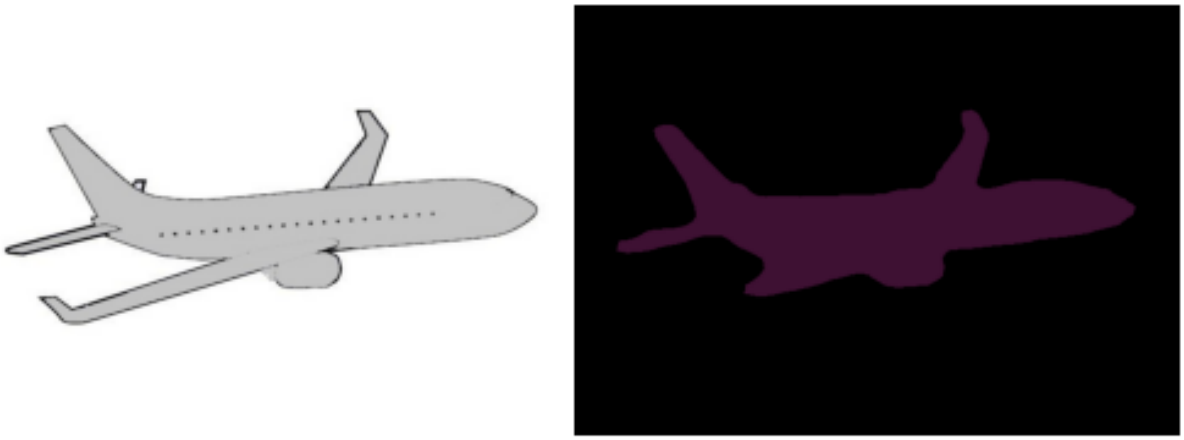
Η απόδοση του DeepLabV3 σε εικόνες που περιείχαν αντικείμενα από τις γνωστές κατηγορίες ήταν πολύ καλή, παρουσιάζοντας καλά οπτικά αποτελέσματα. Στην συνέχεια εξετάσαμε την απόδοση του σε εικόνες με περιγράμματα αντικειμένων από τις 20 γνωστές κατηγορίες αντικειμένων και το κατά πόσο το μοντέλο ήταν ικανό να αντιμετωπίσει τα αντικείμενα αυτά.

Η απάντηση σε αυτό ήταν αρνητική καθώς το μοντέλο δεν ήταν σε θέση να αναγνωρίσει τα γνωστά αντικείμενα από τα περιγράμματα τους, ταξινομώντας τα έτσι ως background.



Εικόνα 29: Περίγραμμα σκύλου που δεν αναγνωρίστηκε από το DeepLabV3. Αριστερά: εικόνα εισόδου Δεξιά: αποτέλεσμα κατάτμησης(αναγνώριση ως background)

Ωστόσο, στη συνέχεια εξετάσαμε τις εικόνες των περιγραμμάτων γεμισμένες με χρώμα. Σε αυτήν την περίπτωση το μοντέλο έδινε ικανοποιητικά αποτελέσματα και ήταν σε θέση να αναγνωρίσει τα αντικείμενα. Άρα, καταλήξαμε στο ότι η χωρική ομαλότητα του χρώματος επηρεάζει την απόδοση του μοντέλου.



Εικόνα 30: Περίγραμμα αεροπλάνου γεμισμένο με γκρι χρώμα και η ταξινόμηση του από το DeepLavV3.

4.2.2 Συμπεριφορά του DeepLabV3 σε άγνωστες κατηγορίες αντικειμένων

Καθώς προαναφέραμε, το DeepLabV3 αναλαμβάνει τον χαρακτηρισμό των pixels σε διάφορες κατηγορίες πάνω στις οποίες έχει εκπαιδευτεί, με τα μη εντοπιζόμενα σε αυτές pixels να κατατάσσονται ως background. Μία αξιοσημείωτη παρατήρηση που προέκυψε κατά τη μελέτη του μοντέλου ήταν η συμπεριφορά του σε αντικείμενα που δεν ανήκουν σε κατηγορίες που γνωρίζει, αλλά μοιάζουν με αυτές που έχει εκπαιδευτεί. Για παράδειγμα, ορισμένα ζώα, όπως η ζέβρα, μπορεί να έχουν ομοιότητες με άλλα ζώα όπως το άλογο, το οποίο ανήκει σε γνωστή κατηγορία. Επίσης, η γωνία λήψης της εικόνας παίζει σημαντικό ρόλο καθώς μία αλεπού θα μπορούσε λανθασμένα να ταξινομηθεί ως γάτα.

Κατά τη διάρκεια πειραματικών δοκιμών, παρατηρήθηκε ότι το μοντέλο δεν παρήγαγε σταθερά αποτελέσματα. Σε ορισμένες περιπτώσεις, όπως αυτές με εύκολα διακεκριμένα χαρακτηριστικά, τα αποτελέσματα ήταν ενθαρρυντικά, αλλά σε πιο δύσκολες περιπτώ-

σεις, όπως αυτές της ζέβρας, τότε λόγω της διχρωμίας της τα αποτελέσματα δεν ήταν ικανοποιητικά.

Ως συμπέρασμα, το μοντέλο φάνηκε να παράγει αποδεκτά αποτελέσματα μόνο για τις γνωστές κατηγορίες. Παρόλα αυτά, η προσοχή μας στράφηκε στο ερώτημα εάν θα μπορούσαμε να αξιοποιήσουμε χαρακτηριστικά(features) του μοντέλου DeepLabV3 για την επίτευξη παραδοσιακής, μη σημασιολογικής κατάτμησης εικόνων.

4.2.3 Εξαγωγή χαρακτηριστικών από το DeepLabV3 για κατάτμηση εικόνας

Το γεγονός ότι το DeepLab είναι ικανό να αναγνωρίσει αντικείμενα μόνο από τις 20 κατηγορίες που έχει εκπαιδευτεί, οδήγησε στο ερώτημα του αν θα μπορούσαμε να χρησιμοποιήσουμε features από κάποιο επίπεδο του DeepLab προκειμένου να κάνουμε κατάτμηση σε εικόνες με επιπλέον αντικείμενα που δεν ανήκουν στις γνωστές κατηγορίες. Έτσι, εξάγοντας τα features από το DeepLab θα μπορούσαμε να τα εκμεταλλευτούμε και να τα χρησιμοποιήσουμε σε συνδυασμό με κάποια άλλη τεχνική ομαδοποίησης ώστε να επιτύχουμε καλύτερα αποτελέσματα. Στην εργασία αυτή ο αλγόριθμος ομαδοποίησης που χρησιμοποιήσαμε ήταν ο UniForCE[10].

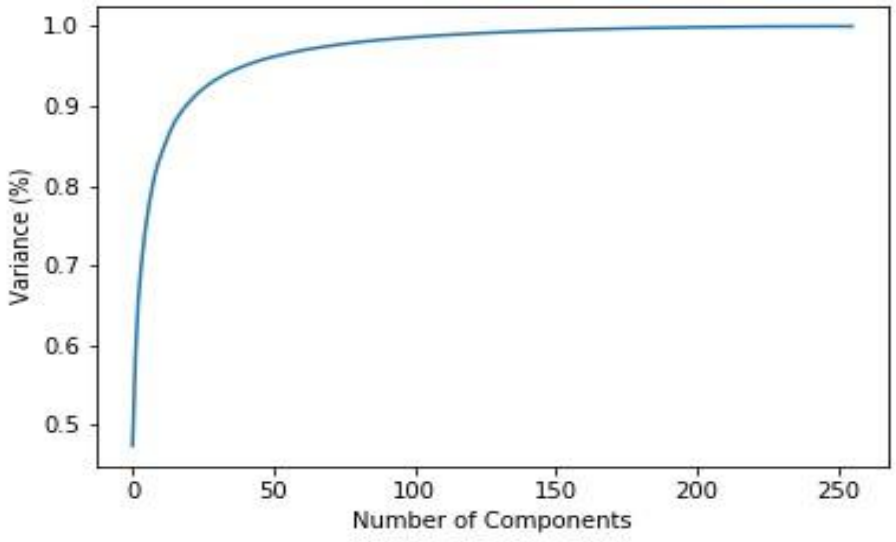
Το κατάλληλο επίπεδο για την εξαγωγή των features είναι το επίπεδο πριν τις τελικές προβλέψεις, δηλαδή το Atrous Spatial Pyramid Pooling(ASPP), το οποίο περιέχει την μέγιστη δυνατή πληροφορία για τις εικόνες. Οι χάρτες χαρακτηριστικών(feature maps) για κάθε pixel στο επίπεδο αυτό αποτελούνται από 256 τιμές. Λόγω της μεγάλης διάστασης των δεδομένων χρησιμοποιήσαμε την τεχνική **PCA** για την μείωση της διάστασης τους και στη συνέχεια τα τροφοδοτούσαμε συνενώνοντας τα με άλλες πληροφορίες π.χ. χρώμα, θέση στον αλγόριθμο ομαδοποίησης.

4.2.3.1 PCA

Η τεχνική Principal Component Analysis(PCA) είναι μια τεχνική γραμμικής μείωσης διαστάσεων με εφαρμογές στην ανάλυση δεδομένων, την οπτικοποίηση και την προεπεξεργασία δεδομένων. Η PCA αποτελεί μία μέθοδο συμπίεσης δεδομένων η οποία συνίσταται από τον επαναπροσδιορισμό των συντεταγμένων ενός συνόλου δεδομένων σε ένα άλλο σύστημα συντεταγμένων το οποίο θα είναι καταλληλότερο για την επικείμενη ανάλυση δεδομένων. Αυτές οι νέες συντεταγμένες είναι το αποτέλεσμα ενός γραμμικού συνδυασμού προερχόμενου από τις αρχικές μεταβλητές και εκπροσωπούνται σε

ορθογώνιο άξονα, ενώ τα επικείμενα σημεία διατηρούν μια φθίνουσα σειρά όσο αφορά στη τιμή της διακύμανσής τους.

Αρχικά στο πρόβλημα μας κατασκευάσαμε ένα γράφημα που εκτιμά την αθροιστική διακύμανση ως προς τον αριθμό των συνιστωσών, με σκοπό την επιλογή του κατάλληλου αριθμού συνιστωσών.



Εικόνα 31: Γράφημα αθροιστικής διακύμανσης για αριθμό συνιστωσών από 0 έως 256.

Από την εικόνα 31 και ύστερα από πειραματισμούς καταλήξαμε ότι 2-5 συνιστώσες εξέφραζαν ικανοποιητική διακύμανση και ήταν αρκετές για να έχουμε καλά αποτελέσματα.

Επομένως, αφού κατεβάζαμε τη διάσταση των δεδομένων με την παραπάνω διαδικασία από 256×H×W σε 2×H×W, όπου H:ύψος της εικόνας και W:πλάτος της εικόνας, τροφοδοτούσαμε τα δεδομένα αυτά συνδυάζοντας την χρωματική πληροφορία των καναλιών LAB της εικόνας στον αλγόριθμο ομαδοποίησης.

Κεφάλαιο 5. UniForCE

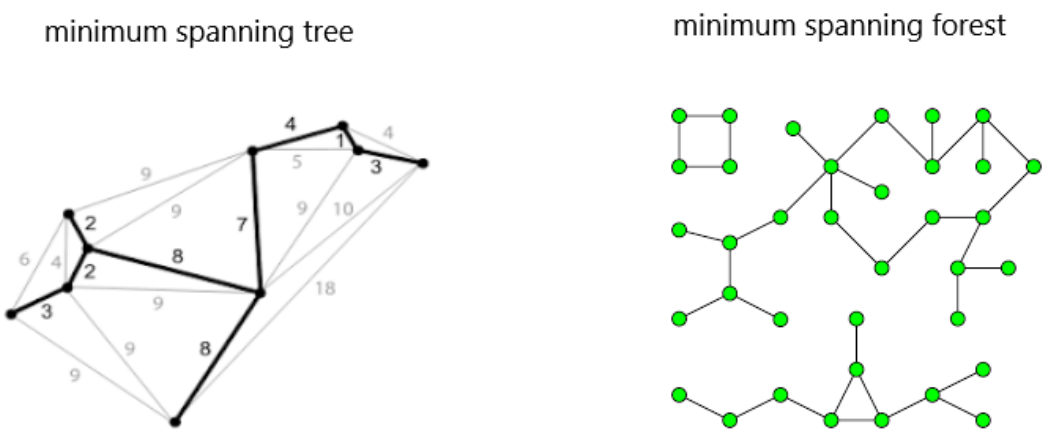
5.1 Αλγόριθμος ομαδοποίησης UniForCE

Ο Unimodality Forest for Clustering Estimation(UniForCE)[10] αποτελεί έναν αλγόριθμο ομαδοποίησης βασισμένο στην ιδέα της μονοτροπικότητας(unimodality), δηλαδή της κατανομής πιθανότητας που η γραφική της παράσταση παρουσιάζει μία μόνο διακριτή κορυφή. Παράλληλα με την ομαδοποίηση ο αλγόριθμος υπολογίζει τον αριθμό των ομάδων k της ομαδοποίησης χωρίς την αρχική γνώση αυτού. Στην εργασία[10] προτείνεται ένας ευέλικτος ορισμός συστάδων(clusters) που ονομάζονται **locally unimodal clusters.** Ένα τέτοιο cluster μπορεί να ληφθεί συναθροίζοντας subclusters από μια αρχική διαμέριση overclustering μέσω μιας διαδικασίας συγχώνευσης που εκτείνεται για όσο διάστημα διατηρείται τοπικά η μονοτροπικότητα στα ζεύγη των subclusters. Για να εξεταστεί αυτή η τοπική ιδιότητα, προτείνεται μια αποτελεσματική στατιστική προσέγγιση που ονομάζεται δοκιμή μονοτροπικού ζεύγους(**unimodal pair testing**) και βασίζεται στο dip-test της μονοτροπικότητας(Dip Test of Unimodality)[11].

5.1.1 Προαπαιτούμενα

Στο σημείο αυτό είναι σημαντικό να αναφέρουμε τους ορισμούς βασικών όρων που θα χρειαστούν στη συνέχεια:

Minimum spanning forest(Ελάχιστο σκελετικό δάσος): Αρχικά, στην θεωρία γραφημάτων, ένα minimum spanning tree είναι ένα υποσύνολο των ακμών ενός συνδεδεμένου, μη κατευθυνόμενου γραφήματος *G(V, E)* με βάρη στις ακμές του που ενώνει όλες τις κορυφές του *G* μαζί, χωρίς την παρουσία κύκλων και έχοντας τον ελάχιστο δυνατό άθροισμα βαρών. Γενικότερα, κάθε μη κατευθυνόμενο γράφημα(όχι απαραίτητα συνδεδεμένο) με βάρη ακμών έχει ένα minimum **spanning forest**, το οποίο είναι μια ένωση των minimum spanning trees των συνιστωσών(components) του.



Εικόνα 33: Στην Αριστερά εικόνα φαίνεται το minimum spanning tree ενός γραφήματος. Στην δεξιά εικόνα βλέπουμε ένα γράφημα που αποτελείται από 3 συνιστώσες, άρα το minimum spanning forest του αποτελείται από 3 minimum spanning trees(ένα για κάθε συνιστώσα).

Ο υπολογισμός του minimum spanning forest στον αλγόριθμο του UniForCE γίνεται χρησιμοποιώντας τον αλγόριθμο του Kruskal[12].

Unimodality(Μονοτροπικότητα): Η μονοτροπικότητα είναι μια στατιστική ιδιότητα που χαρακτηρίζει μια συνάρτηση πυκνότητας πιθανότητας. Μία μονομεταβλητή πυκνότητα πιθανότητας *f* είναι μονοτροπική εάν υπάρχει ένα σημείο $m \in R$, έτσι ώστε η *f* να μην μειώνεται στο $(-\infty,m)$ και να μην αυξάνεται στο (m,∞). Ποιοτικά, αυτό σημαίνει ότι η *f* δέχεται τη μέγιστη τιμή της στο m και μπορεί μόνο να παραμείνει ίδια ή να μειώνεται καθώς απομακρυνόμαστε από αυτό.

Locally Unimodal Cluster(Τοπικά μονοτροπική ομάδα): Εκμεταλλευόμενος την αρχή της μονοτροπικότητας ο UniForCE εισάγει την έννοια του τοπικά μονοτροπικού cluster. Ένα υποσύνολο δεδομένων C ενός dataset X, $C \subseteq X$, είναι τοπικά μονοτροπικό cluster, αν υπάρχει μία διαμέριση $C^{+}= \{c_1, ..., c_K\}$ του C σε subclusters τέτοια ώστε για κάθε ζεύγος (c_i , c_j), να υπάρχει μία ακολουθία S_{ij} των διακριτών subclusters, $S_{ij} = \{s_1 = c_i, s_1, ..., s_{n-1}, s_n = c_j\}$, όπου η ένωση δύο διαδοχικών subclusters $s_i \cup s_{i+1}$ να είναι μονοτροπική, δηλαδή να διατηρεί την μονοτροπικότητα.

Έτσι, καταλήγουμε ότι μία ομαδοποίηση C του dataset X είναι τοπικά μονοτροπική ομαδοποίηση εάν κάθε cluster στο C είναι τοπικά μονοτροπικό.

5.1.2 Βήματα του αλγορίθμου

Ο αλγόριθμος αποτελείται από τα εξής βήματα:

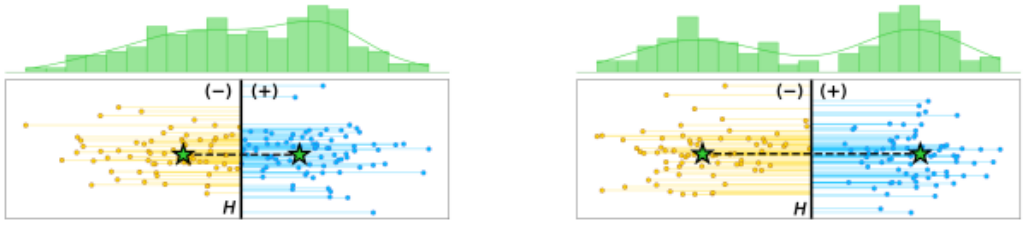
1. Overclustering: Αρχική τοποθέτηση στο σύνολο δεδομένων πολύ περισσότερων subclusters(μικρά σε μέγεθος clusters που απαρτίζονται από λίγα σημεία) από τον πραγματικό αριθμό clusters που υπάρχουν.
2. Unimodal pair testing: Εφαρμογή στατιστικής δοκιμής για τον προσδιορισμό του αν μία ένωση δύο subclusters διατηρεί την μονοτροπικότητα.
3. Ομαδοποίηση με συνάθροιση subclusters: Υπολογισμός των τελικών clusters προσδιορίζοντας τις συνιστώσες του γραφήματος μονοτροπικότητας.

5.1.2.1 Overclustering

Στο overclustering, για τον υπολογισμό των subclusters χρησιμοποιείται κάποιος κλασικός αλγόριθμος ομαδοποίησης όπως ο k-means++ ή ο global k-means++[14], και έχοντας ως παράμετρο **k** μία μεγάλη τιμή π.χ. 50 πετυχαίνετε η υπερομαδοποίηση στο dataset. Στο επόμενο βήμα, αφού έχουν οριστεί τα subclusters, ο αλγόριθμος χρησιμοποιεί μία διαδικασία προεπεξεργασίας για την αφαίρεση των subclusters που περιέχουν πολύ λίγα σημεία. Έτσι, σημεία που μένουν χωρίς ομάδα πλέον τοποθετούνται στα εναπομείναντα clusters σύμφωνα με το βήμα ανάθεσης ομάδων του k-means.

5.1.2.2 Unimodal pair testing

Μετά το βήμα του overclustering, ακολουθεί η διαδικασία της ένωσης subclusters. Στο βήμα αυτό αποφασίζεται αν η ένωση δύο subclusters οδηγεί σε cluster που διατηρεί την μονοτροπικότητα. Έστω μ_i, μ_j τα κέντρα δύο subclusters. Αρχικά, ορίζουμε το διάνυσμα $r_{ij} = \mu_i - \mu_j$ που συνδέει αυτά τα κέντρα, καθώς και το κάθετο υπερεπίπεδο διχοτόμησης H_{ij} του r_{ij} που διέρχεται από το μέσο του. Στη συνέχεια, θεωρούμε το μονομεταβλητό σύνολο P_{ij} που περιέχει την προσημασμένη απόσταση κάθε σημείου c_{ij} από το υπερεπίπεδο H_{ij}. Για τον υπολογισμό της μονοτροπικότητας εφαρμόζεται ο αλγόριθμος dip-test[11] στο σύνολο P_{ij} με μία παράμετρο στατιστικής σημαντικότητας **α**, όπου αποφασίζεται αν το αποτέλεσμα της ένωσης θα είναι μονοτροπικό.

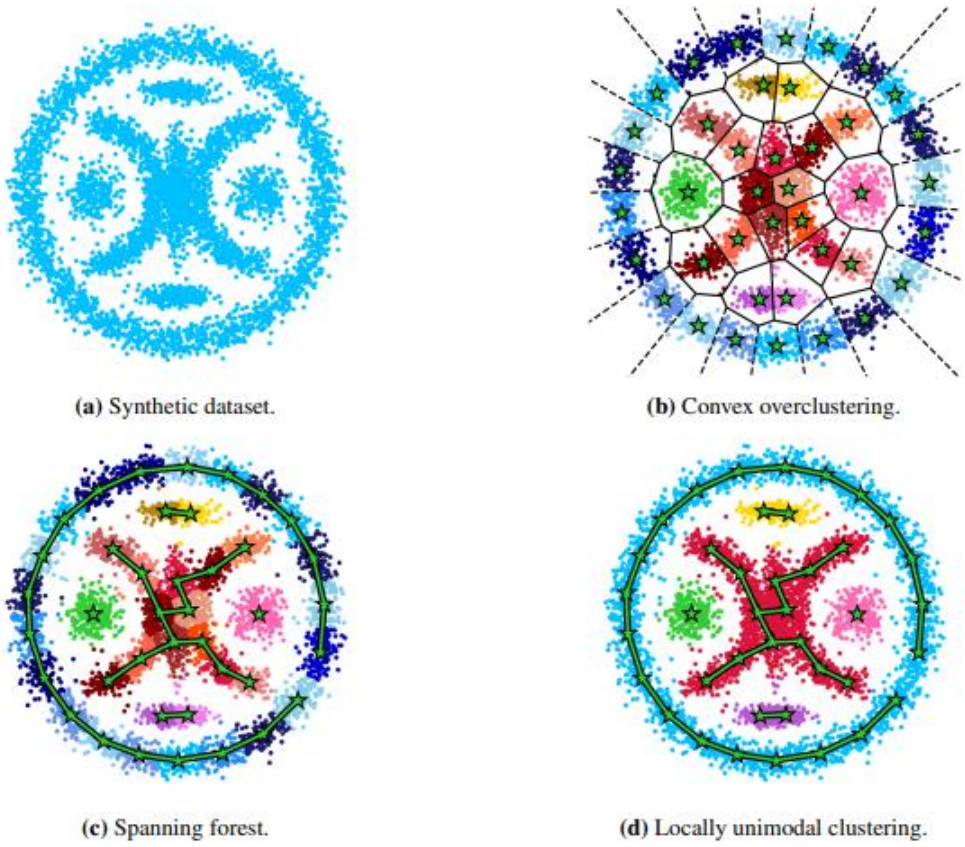


(a) Unimodal case: Two clusters forming a unimodal pair. **(b) Multimodal case:** Two clusters forming a multimodal pair.

Εικόνα 34: Παράδειγμα Unimodal pair testing: Δύο subclusters, ci και cj , εμφανίζονται με κίτρινο και μπλε, αντίστοιχα, και τα κέντρα τους εμφανίζονται ως αστέρια. Η διακεκομμένη γραμμή συνδέει τα δύο κέντρα, ενώ η κάθετη γραμμή είναι η διχοτόμηση του υπερεπίπεδου Hij. Τα ιστογράμματα τους παρουσιάζουν την πυκνότητα του συνόλου Pij, με τις προσημασμένες αποστάσεις από σημείο προς υπερεπίπεδο, τις οποίες δοκιμάζουμε για μονοτροπικότητα χρησιμοποιώντας το dip-test. α) Μονοτροπική περίπτωση: Δεν παρατηρείται χάσμα στις πυκνότητες μεταξύ των subclusters και το Pij κρίνεται ως μονοτροπικό. β) Πολυτροπική περίπτωση: Παρατηρείται ένα σημαντικό χάσμα πυκνότητας μεταξύ των subclusters και το Pij αποφασίζεται ως πολυτροπικό.

5.1.2.3 Τελική ομαδοποίηση

Η εφαρμογή του unimodal pair testing στα ζεύγη των subclusters παρέχει το γράφημα μονοτροπικότητας του αρχικού overclustering. Το επόμενο βήμα είναι η εύρεση των συνιστωσών του γραφήματος μονοτροπικότητας και η τελική εύρεση των clusters. Στη φάση αυτή ο αλγόριθμος εκτελώντας μία παραλλαγή του αλγορίθμου του Kruskal[12] αναπαριστά τα clusters ως ένα σκελετικό δάσος(spanning forest) όπου κάθε σκελετικό του δένδρο αποτελεί ένα locally unimodal cluster. Έτσι, το σύνολο των locally unimodal clusters που θα βρεθούν θα απαρτίζουν τα τελικά clusters της ομαδοποίησης.



Εικόνα 32: Εικόνες των βημάτων που ακολουθούνται από την μεθοδολογία ομαδοποίησης UniForCE σε ένα σύνθετο σύνολο δεδομένων.

Στην πρώτη εικόνα(a) φαίνεται το σύνολο δεδομένων προς ομαδοποίηση. Στην εικόνα(b) παρουσιάζεται το overclustering του συνόλου δεδομένων σε μεγάλο αριθμό ομογενών υποσυστάδων που βρίσκονται σε κυρτές περιοχές. Έπειτα, στην εικόνα (c) βλέπουμε τον υπολογισμό του ελάχιστου σκελετικού δάσους(minimum spanning forest) με βάση τα ζεύγη των subclusters που είναι από κοινού μονοτροπικά (μονοτροπικά ζεύγη). Οι αποσυνδεμένες συνιστώσες του δάσους παρέχουν τα τοπικά μονοτροπικά clusters(εικόνα d).

Κεφάλαιο 6. Αλγόριθμοι κατάτμησης

6.1 Αλγόριθμος κατάτμησης με χρήση features από το προ-εκπαιδευμένο DeepLabV3

Με βάση την ανάλυση όλων των πληροφοριών που παρουσιάστηκαν στο **3.2,** παρουσιάζεται στην εργασία [13] ένας αλγόριθμος κατάτμησης εικόνων που εκμεταλλεύεται τα ενδιάμεσα χαρακτηριστικά του επιπέδου ASPP του δικτύου DeepLabV3 και τις χρωματικές πληροφορίες από το LAB color space.

Περιγραφή αλγορίθμου: [13]

1. Παίρνουμε σαν είσοδο μια RGB εικόνα διαστάσεων H×W×3.
2. Εισάγουμε την εικόνα στο προ-εκπαιδευμένο DeepLabV3 και εξάγουμε τα features από το στρώμα μετά τη χωρική πυραμίδα ASPP του δικτύου. Το σύνολο δεδομένων που επεξεργαζόμαστε πλέον είναι μεγέθους H×W×256.
3. Με τη βοήθεια της μεθόδου PCA μειώνουμε τη διάσταση των δεδομένων που επεξεργάζεται ο αλγόριθμος από H×W×256 σε H×W×2.
4. Μετατρέπουμε την αρχική είσοδο από RGB στον LAB χρωματικό χώρο, παίρνοντας έναν πίνακα διαστάσεων H×W×3 και προσθέτουμε αυτή την πληροφορία στα δεδομένα του βήματος 3, στοιβάζοντας έτσι την πληροφορία του χρώματος για κάθε pixel μαζί με την πληροφορία των PCA features. Άρα η διάσταση των δεδομένων μετατρέπεται από H×W×2 σε H×W×5. Ουσιαστικά μετά από αυτό το βήμα κάθε pixel της εικόνας περιγράφεται από 2 τιμές PCA DeepLabV3 features και 3 τιμές LAB color space.

5. Εισάγουμε την αρχική εικόνα στο προ-εκπαιδευμένο DeepLabV3 και παίρνουμε τις τελικές προβλέψεις για όλα τα pixels και στη συνέχεια πραγματοποιούμε τον ακόλουθο έλεγχο:

i) Αν το δίκτυο δεν αναγνωρίσει κανένα αντικείμενο στην εικόνα τότε τροποποιούμε τις 5 τιμές που περιγράφουν τα pixels του αλγορίθμου μας με τον εξής τρόπο: τις 2 τιμές PCA features τις πολλαπλασιάζουμε με βάρος w=0,2 και τις 3 τιμές LAB τις πολλαπλασιάζουμε με βάρος 1-w=0,8. ***ii)*** Αν το δίκτυο αναγνωρίζει τουλάχιστον ένα αντικείμενο στην εικόνα, τότε επαναληπτικά για όλα τα pixels γίνεται ο εξής έλεγχος:

- Αν το τρέχον pixel δεν αναγνωρίζεται από το δίκτυο (άρα ταξινομείται ως background) τότε τροποποιούμε τις 5 τιμές που περιγράφουν το αντίστοιχο pixel του αλγορίθμου μας με τον εξής τρόπο: τις 2 τιμές PCA features τις πολλαπλασιάζουμε με βάρος w=0,1 και τις 3 τιμές LAB τις πολλαπλασιάζουμε με βάρος 1-w=0,9.
- Αν το τρέχον pixel αναγνωρίζεται από το δίκτυο τότε θέτουμε και τις 5 τιμές που περιγράφουν το αντίστοιχο pixel του αλγορίθμου μας ίσες με 0.

Τα βάρη καθορίζουν την σημαντικότητα των τιμών features και LAB και επιλέχθηκαν μετά από πολλά πειράματα καθώς κρίθηκαν αποτελεσματικές ως προς τα αποτελέσματα που παρουσίαζε ο αλγόριθμος σε μεγάλο εύρος εικόνων.

6. Με τη βοήθεια του silhouette score επιλέγουμε τον επιθυμητό αριθμό από ομάδες k που θα χρησιμοποιήσουμε, αγνοώντας την τιμή 2 και ελέγχοντας μόνο μικρές σχετικά τιμές k (δηλαδή $3 \leq k \leq 6$).

7. Εκτελούμε τον αλγόριθμο k-means με είσοδο τα δεδομένα που έχουν προκύψει από το βήμα 5 και το k που έχουμε επιλέξει από το βήμα 6.

8. Σαν έξοδο παίρνουμε την κατάτμηση της αρχικής εικόνας με βάση τις ομάδες που υπολογίζονται στο βήμα 7.

Συνοψίζοντας, η προαναφερθείσα μέθοδος πραγματοποιεί κατάτμηση εικόνων αξιοποιώντας την τελική γνώση του δικτύου DeepLabV3 για τον εντοπισμό των γνωστών στο δίκτυο αντικειμένων και ταυτόχρονα χρησιμοποιεί έναν σταθμισμένο συνδυασμό ενδιάμεσων features και χρωματικής (LAB) πληροφορίας για να πραγματοποιήσει μια ομαλή ομαδοποίηση χωρίς θόρυβο των pixels που δεν ανήκουν σε κάποια από τις 20 γνωστές κατηγορίες. Επιπλέον, ο αλγόριθμος λειτουργεί τελείως αυτόματα, αφού ο υπολογισμός των ομάδων στο βήμα 6 υπολογίζεται αυτόματα μέσω του silhouette score και έτσι δεν απαιτείται από τον χρήστη η είσοδος κάποιας παραμέτρου.

Εικόνα 33: Παραδείγματα κατάτμησης εικόνων από τον αλγόριθμο[13]. (Αριστερή στήλη) Αρχικές εικόνες, (Μεσαία στήλη) Έξοδος του DeepLabV3, (δεξιά στήλη) Κατάτμηση της εικόνας μέσω του αλγορίθμου[13].

Από τις παραπάνω εικόνες είναι εμφανής η δυνατότητα κατάτμησης αντικειμένων που δεν ανήκουν στις γνωστές κατηγορίες που έχει εκπαιδευτεί το DeepLabV3.

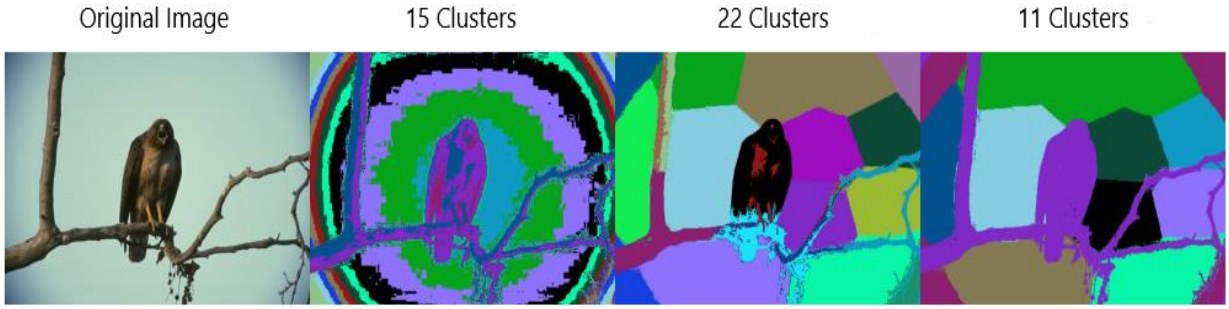
6.2 Αλγόριθμος κατάτμησης με χρήση του UniForCE και features από το DeepLabV3

Στην παρούσα εργασία μελετήθηκε η χρήση του αλγορίθμου ομαδοποίησης UniForCE που παρουσιάστηκε **κεφάλαιο 5** ως μέθοδος για κατάτμηση εικόνας. Πιο συγκεκριμένα, προσθέσαμε τον αλγόριθμο του UniForCE ως μέθοδο ομαδοποίησης στο βήμα 7 της διαδικασίας κατάτμησης με χρήση deep features που προτάθηκε από την εργασία[13].

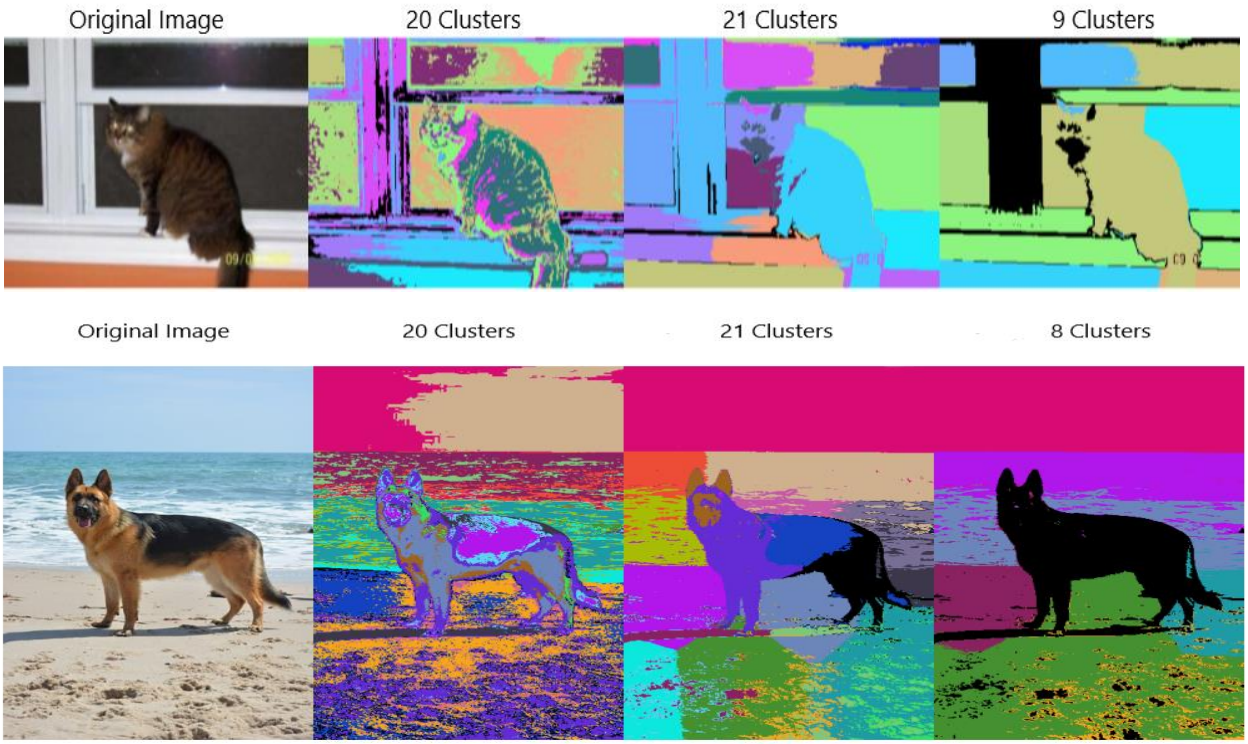
Έτσι, ύστερα από πολλά πειράματα, προτείνουμε έναν αλγόριθμο κατάτμησης εικόνας που χρησιμοποιεί την τελική κατάτμηση του DeepLabV3, τα deep features του επιπέδου Atrous Spatial Pyramid Pooling(ASPP), τις LAB τιμές χρώματος και την χωρική πληροφορία των θέσεων των pixels.

6.2.1 Κατάτμηση εικόνας με UniForCE

Σε αρχική φάση εξετάσαμε την απόδοση του UniForCE για την κατάτμηση εικόνας σε περιοχές χρησιμοποιώντας μόνο τις τρεις τιμές του χρώματος RGB της εικόνας. Ωστόσο, το γεγονός ότι απουσίαζε η χωρική πληροφορία από τα δεδομένα προς ομαδοποίηση, έκανε την κατάτμηση πολύ ευαίσθητη στο χρώμα, πράγμα που οδηγούσε σε πολύ μεγάλο αριθμό ομάδων. Έτσι, αποφασίσαμε την εισαγωγή και της πληροφορίας θέσης για το κάθε pixel εκτελώντας δύο φορές τον αλγόριθμο UniForCE. Κατά την πρώτη εκτέλεση χρησιμοποιούσαμε τις τρεις τιμές RGB και τις δύο τιμές της θέσης (x, y), δηλαδή κάθε pixel αναπαρίστανται από 5 τιμές. Από το αποτέλεσμα της ομαδοποίησης της πρώτης εκτέλεσης, αποθηκεύαμε τα κέντρα των clusters και σε ποιο cluster ανήκει κάθε pixel, δηλαδή τις ετικέτες/labels. Στη συνέχεια, εκτελούσαμε ξανά τον UniForCE με είσοδο μόνο τις τιμές RGB και επιστρέφοντας στο βήμα του overclustering ως subclusters τα κέντρα των clusters και τα labels του κάθε pixel που υπολογίστηκαν κατά την πρώτη εκτέλεση. Με αυτόν τρόπο καταφέρναμε να έχουμε μια πιο ομαλή κατάτμηση των περιοχών, όπως φαίνεται παρακάτω:



Εικόνα 34: (1η στήλη) Αρχική εικόνα, (2η στήλη) Κατάτμηση εικόνας με χρήση των RGB παραμέτρων, (3η στήλη) Κατάτμηση εικόνας με χρήση RGB και της θέσης(x, y) των pixel, (4η στήλη) Αποτέλεσμα κατάτμησης με εκτέλεση του UniForCE δύο φορές όπως αναλύθηκε παραπάνω.



Εικόνα 34: (1η στήλη) Αρχική εικόνα, (2η στήλη) Κατάτμηση εικόνας με χρήση των RGB παραμέτρων, (3η στήλη) Κατάτμηση εικόνας με χρήση RGB και της θέσης(x, y) των pixel, (4η στήλη) Αποτέλεσμα κατάτμησης με εκτέλεση του UniForCE δύο φορές όπως αναλύθηκε παραπάνω.

Στα παραπάνω παραδείγματα οι παράμετροι του UniForCE ήταν: k=30 subclusters, $\alpha=10^{-5}$ στατιστική σημαντικότητα του dip-test.

6.2.2 Κατάτμηση με χρήση των features του DeepLabV3

Σε δεύτερη φάση, εξετάστηκε η χρήση του UniForCE ως μέθοδος ομαδοποίησης στο βήμα 7 της διαδικασίας κατάτμησης με χρήση deep features που προτάθηκε από την εργασία[13], καθώς με αυτόν τον τρόπο θα μπορούσαν να αναγνωριστούν αντικείμενα των 20 γνωστών κατηγοριών που έχει εκπαιδευτεί ο DeepLabV3. Λόγω της δυνατότητας του UniForCE να προσδιορίζει τον αριθμό τον ομάδων κατά την εκτέλεση του, αφαιρέθηκε η χρήση του silhouette score. Αρχικά, εξετάσαμε την λειτουργία του αλγορίθμου αλλάζοντας μόνο τον αλγόριθμο ομαδοποίησης από k-means σε UniForCE. Ωστόσο, ενώ τα αντικείμενα που αναγνωριζόταν από το DeepLabV3 διατηρούνταν στην τελική κατάτμηση, ο αλγόριθμος υπερομαδοποιούσε την υπόλοιπη εικόνα. Όπως και παραπάνω (ενότητα 6.2.1) για την αντιμετώπιση αυτού του φαινομένου αποφασίσαμε την διπλή εκτέλεση του αλγορίθμου UniForCE. Έτσι, κατά την πρώτη εκτέλεση για κάθε pixel χρησιμοποιούσαμε τις 2 τιμές των PCA DeepLabV3 features, τις 3 τιμές για την χρωματική αναπαράσταση LAB και την θέση (x, y) του pixel, δηλαδή κάθε pixel αναπαρίστανται από συνολικά 7 τιμές. Στη συνέχεια, εκτελούσαμε ξανά τον UniForCE αφαι-

ρώντας από τα δεδομένα τις τιμές της θέσης των pixels και επιστρέφοντας στο βήμα του overclustering ως subclusters τα κέντρα των clusters και τα labels των pixels που υπολογίστηκαν κατά την πρώτη εκτέλεση. Έτσι, πετυχαίναμε μια πιο ομαλή κατάτμηση των περιοχών της εικόνας.

Περιγραφή αλγορίθμου(Double UniForCE):

1. Παίρνουμε σαν είσοδο μια RGB εικόνα διαστάσεων H×W×3.

2. Εισάγουμε την εικόνα στο προ-εκπαιδευμένο DeepLabV3 και εξάγουμε τα features από το στρώμα μετά τη χωρική πυραμίδα ASPP του δικτύου. Το σύνολο δεδομένων που επεξεργαζόμαστε πλέον είναι μεγέθους H×W×256.

3. Με τη βοήθεια της μεθόδου PCA μειώνουμε τη διάσταση των δεδομένων που επεξεργάζεται ο αλγόριθμος από H×W×256 σε H×W×2.

4. Μετατρέπουμε την αρχική είσοδο από RGB στον LAB χρωματικό χώρο, παίρνοντας έναν πίνακα διαστάσεων H×W×3 και προσθέτουμε αυτή την πληροφορία στα δεδομένα του βήματος 3, στοιβάζοντας έτσι την πληροφορία του χρώματος για κάθε pixel μαζί με την πληροφορία των PCA features. Άρα η διάσταση των δεδομένων μετατρέπεται από H×W×2 σε H×W×5. Ουσιαστικά μετά από αυτό το βήμα κάθε pixel της εικόνας περιγράφεται από 2 τιμές PCA DeepLabV3 features και 3 τιμές LAB color space.

5. Εισάγουμε την αρχική εικόνα στο προ-εκπαιδευμένο DeepLabV3 και παίρνουμε τις τελικές προβλέψεις για όλα τα pixels και στη συνέχεια πραγματοποιούμε τον ακόλουθο έλεγχο:

i) Αν το δίκτυο δεν αναγνωρίσει κανένα αντικείμενο στην εικόνα τότε τροποποιούμε τις 5 τιμές που περιγράφουν τα pixels του αλγορίθμου μας με τον εξής τρόπο: τις 2 τιμές PCA features τις πολλαπλασιάζουμε με βάρος w=0,2 και τις 3 τιμές LAB τις πολλαπλασιάζουμε με βάρος 1-w=0,8. ***ii)*** Αν το δίκτυο αναγνωρίζει τουλάχιστον ένα αντικείμενο στην εικόνα, τότε επαναληπτικά για όλα τα pixels γίνεται ο εξής έλεγχος:

- Αν το τρέχον pixel δεν αναγνωρίζεται από το δίκτυο (άρα ταξινομείται ως background) τότε τροποποιούμε τις 5 τιμές που περιγράφουν το αντίστοιχο pixel του αλγορίθμου μας με τον εξής τρόπο: τις 2 τιμές PCA features τις πολλαπλασιάζουμε με βάρος w=0,1 και τις 3 τιμές LAB τις πολλαπλασιάζουμε με βάρος 1-w=0,9.

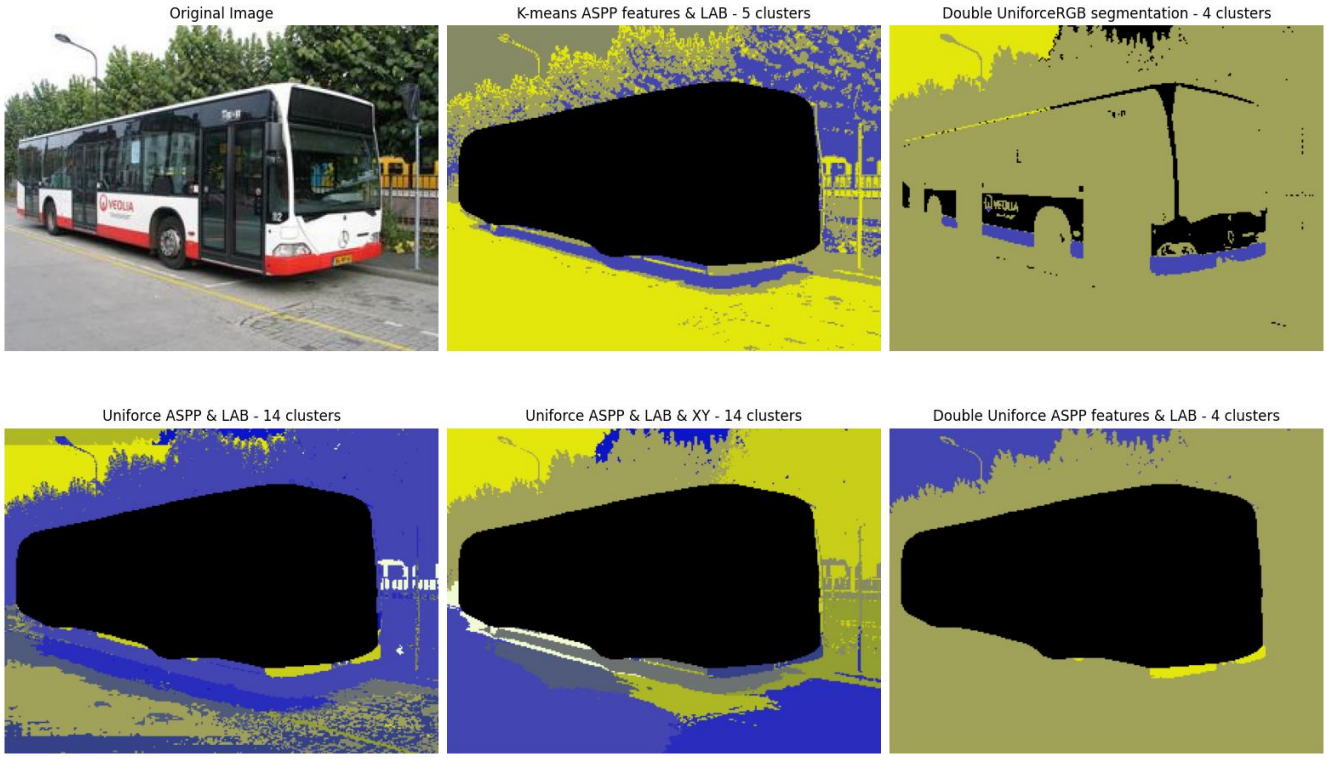
- Αν το τρέχον pixel αναγνωρίζεται από το δίκτυο τότε θέτουμε και τις 5 τιμές που περιγράφουν το αντίστοιχο pixel του αλγορίθμου μας ίσες με 0.

7. Εκτελούμε τον αλγόριθμο UniForCE με είσοδο τα δεδομένα που έχουν προκύψει από το βήμα 5 και τις θέσεις(x, y) των pixels. Αποθηκεύουμε τα κέντρα των ομάδων που υπολογίζονται.

8. Εκτελούμε ξανά τον αλγόριθμο UniForCE με είσοδο μόνο τα δεδομένα που έχουν προκύψει από το βήμα 5, επιστρέφοντας στο βήμα του overclustering ως subclusters τα κέντρα που υπολογίστηκαν στην ομαδοποίηση του βήματος 7.

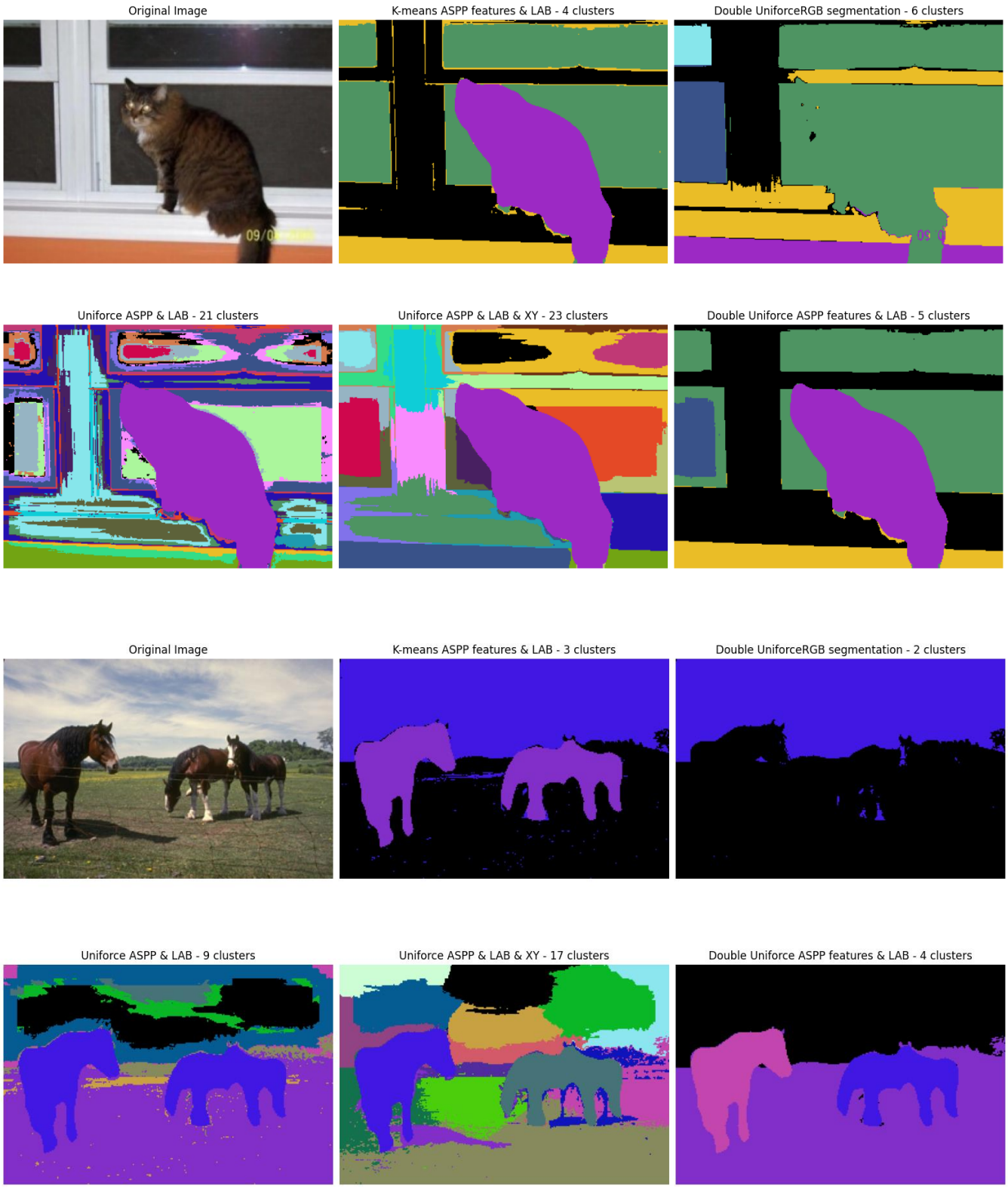
9. Σαν έξοδο παίρνουμε την κατάτμηση της αρχικής εικόνας με βάση τις ομάδες που υπολογίζονται στο βήμα 8.

Εικόνες αποτελεσμάτων:



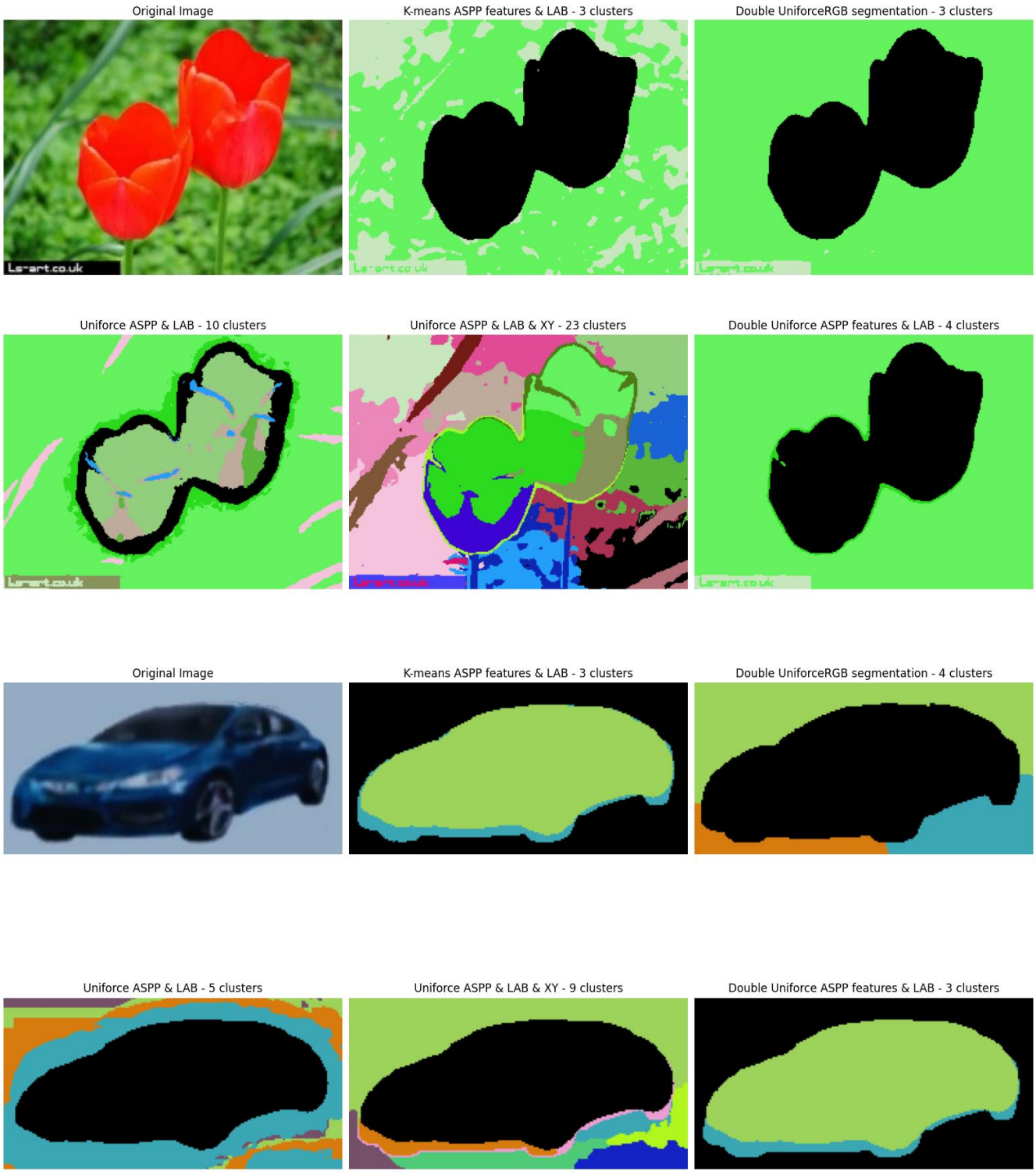
Εικόνα 34: Παραδείγματα κατάτμησης εικόνων. 1η γραμμή: (1η στήλη) Αρχική εικόνα, (2η στήλη) αλγόριθμος κατάτμησης[13] με χρήση k-means, (3η στήλη) διπλή εκτέλεση UniForCE με δεδομένα της τιμές RGB όπως αναλύθηκε στην ενότητα 6.2.1.

2η γραμμή: (1η στήλη) αλγόριθμος κατάτμησης[13] με χρήση UniForCE ως αλγορίθμου ομαδοποίησης, (2η στήλη) παραλλαγή αλγόριθμου κατάτμησης[13] με χρήση UniForCE και εισαγωγή πληροφορίας θέσης(x, y) στα δεδομένα προς ομαδοποίηση, (3η στήλη) αλγόριθμος κατάτμησης(Double UniForCE) με διπλή εκτέλεση του UniForCE.



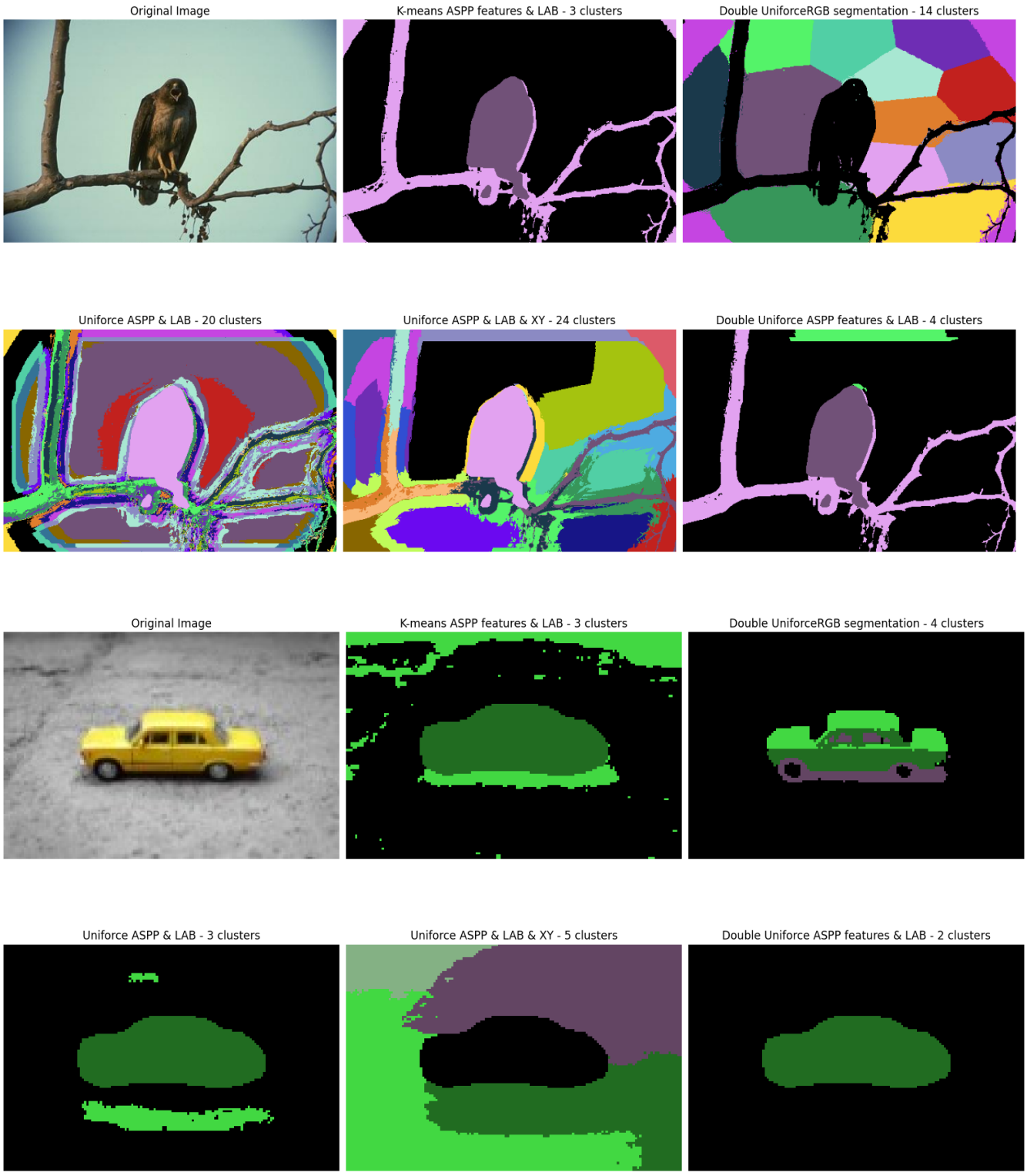
Εικόνα 34: Παραδείγματα κατάτμησης εικόνων. 1η γραμμή: (1η στήλη) Αρχική εικόνα, (2η στήλη) αλγόριθμος κατάτμησης[13] με χρήση k-means, (3η στήλη) διπλή εκτέλεση UniForCE με δεδομένα της τιμές RGB όπως αναλύθηκε στην ενότητα 6.2.1.

2η γραμμή: (1η στήλη) αλγόριθμος κατάτμησης[13] με χρήση UniForCE ως αλγορίθμου ομαδοποίησης, (2η στήλη) παραλλαγή αλγόριθμου κατάτμησης[13] με χρήση UniForCE και εισαγωγή πληροφορίας θέσης(x, y) στα δεδομένα προς ομαδοποίηση, (3η στήλη) αλγόριθμος κατάτμησης(Double UniForCE) με διπλή εκτέλεση του UniForCE.



Εικόνα 34: Παραδείγματα κατάτμησης εικόνων. 1η γραμμή: (1η στήλη) Αρχική εικόνα, (2η στήλη) αλγόριθμος κατάτμησης[13] με χρήση k-means, (3η στήλη) διπλή εκτέλεση UniForCE με δεδομένα της τιμές RGB όπως αναλύθηκε στην ενότητα 6.2.1.

2η γραμμή: (1η στήλη) αλγόριθμος κατάτμησης[13] με χρήση UniForCE ως αλγορίθμου ομαδοποίησης, (2η στήλη) παραλλαγή αλγόριθμου κατάτμησης[13] με χρήση UniForCE και εισαγωγή πληροφορίας θέσης(x, y) στα δεδομένα προς ομαδοποίηση, (3η στήλη) αλγόριθμος κατάτμησης(Double UniForCE) με διπλή εκτέλεση του UniForCE.



Εικόνα 34: Παραδείγματα κατάτμησης εικόνων. 1η γραμμή: (1η στήλη) Αρχική εικόνα, (2η στήλη) αλγόριθμος κατάτμησης[13] με χρήση k-means, (3η στήλη) διπλή εκτέλεση UniForCE με δεδομένα της τιμές RGB όπως αναλύθηκε στην ενότητα 6.2.1.

2η γραμμή: (1η στήλη) αλγόριθμος κατάτμησης[13] με χρήση UniForCE ως αλγορίθμου ομαδοποίησης, (2η στήλη) παραλλαγή αλγόριθμου κατάτμησης[13] με χρήση UniForCE και εισαγωγή πληροφορίας θέσης(x, y) στα δεδομένα προς ομαδοποίηση, (3η στήλη) αλγόριθμος κατάτμησης(Double UniForCE) με διπλή εκτέλεση του UniForCE.

Σε όλα τα παραπάνω παραδείγματα οι παράμετροι του UniForCE ήταν: k=30 subclusters, $\alpha=10^{-5}$ στατιστική σημαντικότητα του dip-test.

Κεφάλαιο 7. Συμπεράσματα

Με βάση τα παραπάνω παραδείγματα κατάτμησης εικόνων και ύστερα από πολλά πειράματα καταλήξαμε ότι η χρήση του αλγορίθμου Double UniForCE συνδυάζοντας βαθιά χαρακτηριστικά(deep features) του μοντέλου DeepLabV3 παρουσιάζει πολύ καλά αποτελέσματα κατάτμησης εικόνας. Το γεγονός ότι ο UniForCE δεν απαιτεί την αρχική γνώση του αριθμού των ομάδων k, αυξάνει δραματικά την ταχύτητα εκτέλεσης σε σχέση με τη χρήση του silhouette score για τον υπολογισμό των ομάδων που προτάθηκε στον αλγόριθμο της εργασίας [13], με αντάλλαγμα ίσως σε ορισμένες περιπτώσεις την εισαγωγή κάποιων λίγων περιττών clusters στο background της κατατμημένης εικόνας. Τέλος, χάρη στα εντυπωσιακά αποτελέσματα του DeepLabV3 στην αναγνώριση αντικειμένων πάνω στα οποία έχει εκπαιδευτεί προσφέρεται εξαιρετική ακρίβεια στην αναγνώριση και κατάτμηση των γνωστών αντικειμένων, προσθέτοντας ένα επιπλέον επίπεδο αξιοπιστίας στη συνολική διαδικασία κατάτμησης.

Βιβλιογραφία

[1] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[2] Long, Jonathan & Shelhamer, Evan & Darrell, Trevor., Fully Convolutional Networks for Semantic Segmentation, arXiv:1411.4038, 2014.

[3] E. Shelhamer, J. Long and T. Darrell, Fully Convolutional Networks for Semantic Segmentation, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 1 April 2017.

[4] Rousseeuw, P.J., Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics, 20, 53-65, 1987.

[5] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A., The PASCAL visual object classes challenge 2007 (VOC2007) Results, http://www.pascalnetwork.org/ challenges/VOC/voc2007/index.html, 2007.

[6] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Common objects in context, In ECCV, 2014.

[7] K. O'Shea, R. Nash, An Introduction to Convolutional Neural Networks, 2015.

[8] L. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv:1706.05587, 2017.

[9] Arthur, David, and Sergei Vassilvitskii. "k-means++: The advantages of careful seeding." *Soda*. Vol. 7. 2007.

[10] G. Vardakas, A. Likas, A. Kalogeratos, UniForCE: The Unimodality Forest Method for Clustering and Estimation of the Number of Clusters, arXiv:2312.11323v1, 2023.

[11] J. A. Hartigan and P. M. Hartigan, “The Dip Test of Unimodality,” The Annals of Statistics, vol. 13, no. 1, pp. 70 – 84, 1985.

[12] J. B. Kruskal, “On the shortest spanning subtree of a graph and the traveling salesman problem,” Proceedings of the American Mathematical Society, vol. 7, no. 1, pp. 48–50, 1956.

[13] Π. Τσιρώνης, Κατάτμηση Εικόνων με πλήρως συνελικτικά δίκτυα, Διπλωματική Εργασία, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, 2020.

[14] Vardakas, Georgios, and Aristidis Likas. "Global $ k $-means $++ $: an effective relaxation of the global $ k $-means clustering algorithm." *arXiv preprint arXiv:2211.12271* (2022).

[15] Lloyd, Stuart. "Least squares quantization in PCM." *IEEE transactions on information theory* 28.2 (1982): 129-137.