

Θέμα 1 (2 μονάδες)

1.1 Εξετάζουμε τον Λαπλασιανό τελεστή. Για κάθε εικονοστοιχείο υπολογίζει:

- α. την παράγωγο στην κατακόρυφη διεύθυνση
- β. την παράγωγο στην οριζόντια διεύθυνση
- γ. το μέγεθος και την κατεύθυνση των ακμών
- δ. την καμπυλότητα

► Laplacian: τελεστής 2^{ης} τάξης $\nabla^2 f = \partial^2 f / \partial x^2 + \partial^2 f / \partial y^2$

ανιχνεύει σημεία καμπυλότητας, έντονης αλλαγής φωτός - ανεξάρτητος από κατεύθυνση
υψηλή ευαισθησία σε θόρυβο $\Rightarrow$ χρειάζεται ετομάλυνση (low-pass)

ανίχνευση ακμών με zero-crossing

► gradient: τελεστής 1^{ης} τάξης / διάνυσμα $\nabla f = [\partial f / \partial x, \partial f / \partial y]$ μέτρο: $|\nabla f| = [(\partial f / \partial x)^2 + (\partial f / \partial y)^2]^{1/2}$

δείχνει την κατεύθυνση μέγιστης αύξησης φωτεινότητας κατεύθυνση: $\theta = \tan^{-1} (f_y / f_x)$

συνήθως χρησιμοποιούμε Prewitt / Sobel

► Sobel: τελεστής 1^{ης} τάξης, κατεύθυνση και ένταση ακμής + ετομάλυνση συν κάθετη κατεύθ

► Prewitt: όμοιος με Sobel, χωρίς βάρη σε κέντρο, λιγότερο ακριβής σε θόρυβο

1.2 Εξετάστε μια εικόνα στην οποία εφαρμόζεται ο Λαπλασιανός τελεστής για ανίχνευση ακμών. Η ύπαρξη μιας ακμής στην αρχική εικόνα, στην φιλτραρισμένη εικόνα αντιστοιχεί:

- α. σε διάβαση από το μηδέν
- β. σε μηδενισμό
- γ. σε τοπικό μέγιστο
- δ. σε τοπικό ελάχιστο

Laplacian: ανιχνεύει ακμές με αλλαγή προσήμου, δηλαδή διάβαση από μηδέν (zero-crossing)

1.3 Εάν η εικόνα είναι θορυβώδης, τι είδους φίλτρο πρέπει να εφαρμοστεί στην εικόνα πριν από την εφαρμογή του Λαπλασιανού τελεστή;

- α. ένα ανωδιαβατό φίλτρο
- β. ένα κατωδιαβατό φίλτρο
- γ. ένα ζωνοπερατό φίλτρο

Laplacian: ευαίσθητος στον θόρυβο, καλό είναι να προηγείται ετομάλυνση $\sim$ low-pass filters

1.4 Μεταβάλλοντας τη διάσταση αυτού του φίλτρου, μπορούμε να εντοπίσουμε ακμές με:

- α. διαφορετικό contrast
- β. διαφορετικό μήκος
- γ. διαφορετικό resolution

Το kernel size καθορίζει την κλίμακα:

- μεγαλύτερο φίλτρο: εντοπίζει πιο ποτερές, μεγαλύτερες ακμές (π.χ. λογί με $\sigma=5$)
 - μικρότερο φίλτρο: ανιχνεύει λεπτομέρειες, μικρές ακμές (π.χ. λογί με $\sigma=1$)
- } διαφορετικό resolution

Θέμα 2 (2 μονάδες)

2.1 Αυτό εφαρμόζεται σε μια εικόνα grayscale I με pixels των 5-bits. Ποια είναι η μέγιστη και ελάχιστη δυνατή τιμή των pixels της εικόνας J που προκύπτει μετά την εφαρμογή του φίλτρου; Μην κάνετε κανονικοποίηση.

$$H = \begin{bmatrix} 1 & -3 & 0 \\ -3 & 0 & 4 \\ 0 & 4 & 1 \end{bmatrix}$$

$$\text{βαθός } L: 5 \Rightarrow 2^5 = 32 \text{ τιμές} \Rightarrow I(i,j) \in (0, 31)$$

$$J(i,j) = \sum H(m,n) \cdot I(i+m, j+n)$$

$$\begin{aligned} \text{μέγιστη τιμή: } J_{\max} &= 31 \cdot (\sum_{\omega_{\max}} \omega_i) + 0 \cdot (\sum_{\omega_{\min}} \omega_i) = 31 \cdot 10 + 0 \cdot (-6) = 310 \\ \text{ελάχιστη τιμή: } J_{\min} &= 0 \cdot (\sum_{\omega_{\max}} \omega_i) + 31 \cdot (\sum_{\omega_{\min}} \omega_i) = 0 \cdot 10 + 31 \cdot (-6) = -186 \end{aligned} \quad \left. \vphantom{\begin{aligned} J_{\max} &= 31 \cdot (\sum_{\omega_{\max}} \omega_i) + 0 \cdot (\sum_{\omega_{\min}} \omega_i) = 31 \cdot 10 + 0 \cdot (-6) = 310 \\ J_{\min} &= 0 \cdot (\sum_{\omega_{\max}} \omega_i) + 31 \cdot (\sum_{\omega_{\min}} \omega_i) = 0 \cdot 10 + 31 \cdot (-6) = -186 \end{aligned}} \right\} J(i,j) \in (-186, 310)$$

2.2 Ποιος μετασχηματισμός επιπέδου του γκρι πρέπει να εφαρμοστεί στα pixels της J για να φέρει τις τιμές τους πάλι στα 5-bits;

$$J(i,j) \in (-186, 310) \rightarrow J_{\text{scaled}}(i,j) \in (0, 31)$$

$$J_{\text{scaled}}(i,j) = \frac{J(i,j) - J_{\min}}{J_{\max} - J_{\min}} \cdot (2^L - 1) = \frac{J(i,j) + 186}{496} \cdot 31 = \frac{J(i,j) + 186}{16}$$

Θέμα 3 (3 μονάδες)

Εξετάστε μια εφαρμογή βιντεοτηλεφωνίας χρησιμοποιώντας Σύσταση ITU-T H.261. Η ακολουθία βίντεο: Κωδικοποιείται με χωρική ανάλυση CIF (352x288 δείγματα για τη φωτεινότητα, 4:2:0), Ρυθμό καρέ: 10 fps, Κανάλι με σταθερό bitrate: 64 kbit/s, Buffer εξόδου: 12.800 bits.

CIF (352x288 δείγματα Y, 4:2:0): ανάλυση φωτεινότητας (Y), τα χρωματικά κανάλια (C_b, C_r)
frame rate: $R = 10$ fps έχουν μισή ανάλυση ανά διάσταση σε σχέση με Y
bitrate: $B = 64$ kbit/s (CBR: constant bit rate)
output buffer size: 12.800 bits

3.1 Ποιος είναι ο μέσος αριθμός bits κάθε εικόνας που κωδικοποιείται;

$$\text{bits per frame} = \frac{\text{bits}}{\text{sec}} \cdot \frac{\text{sec}}{\text{frame}} = \frac{\text{bitrate}}{\text{framerate}} = \frac{64.000}{10} \frac{\text{bits}}{\text{frame}} = 6400 \text{ bits per frame}$$

3.2 Ποιος είναι ο μέγιστος και ποιος ο ελάχιστος αριθμός bits που μπορούν να χρησιμοποιηθούν για την κωδικοποίηση κάθε εικόνας;

$$\begin{aligned} B_{\max} &= \text{buffer} = 12.800 \text{ bits (όλος ο buffer μπορεί να καταναλωθεί για 1 frame)} \\ B_{\min} &= 0 \text{ (θεωρητικά)} \end{aligned}$$

3.3 Ποιος είναι ο μέγιστος αριθμός bits που μπορεί να χρησιμοποιηθεί για την κωδικοποίηση μιας εικόνας, εάν ο buffer είναι γεμάτος όταν η εικόνα αρχίζει να κωδικοποιείται;

$$\begin{aligned} \text{bits per frame} &= 6400 \text{ bits/frame} \\ B_{\text{avail}} &= 6400 + 12.800 = 19.200 \text{ bits} \end{aligned}$$

3.4 Ποιος είναι ένας τυπικός αριθμός διανυσμάτων κίνησης που μπορούν να χρησιμοποιηθούν για την κωδικοποίηση μιας εικόνας;

Σε πρότυπα συμπίεσης βίντεο, όπως ITU-T H.261, κάθε εικόνα χωρίζεται σε macroblocks μεγέθους 16×16 pixels φωτεινότητας (Y).

Η τεχνική motion compensation εφαρμόζεται στα macroblocks και αντιστοιχεί ένα διάνυσμα κίνησης στο καθένα.

CIF (352×288): $\left\{ \begin{array}{l} 352/16 = 22 \text{ macroblocks (οριζόντια)} \\ 288/16 = 18 \text{ macroblocks (κατακόρυφα)} \end{array} \right\} \Rightarrow 22 \times 18 = 396 \text{ macroblocks} \Rightarrow 396 \text{ διανύσματα κίνησης}$

3.5 Λαμβάνοντας υπόψη ότι χρησιμοποιείται κανάλι σταθερού ρυθμού bits, ποιο τμήμα του κωδικοποιητή λαμβάνει τις πιο σημαντικές αποφάσεις για να επιτρέψει στον κωδικοποιητή να ελέγχει τον αριθμό των bits που δαπανώνται για κάθε εικόνα;

CBR (constant bitrate): ο κωδικοποιητής προσαρμόζει επίπεδο κβαντοποίησης ανά εικόνα.

- πολυπλοκή εικόνα: $\uparrow$ κβαντα $\Rightarrow \downarrow$ bits/μοτίβο

- απλή εικόνα: $\downarrow$ κβαντα $\Rightarrow \uparrow$ bits/μοτίβο

Ελέγχος bit ανά εικόνα σε CBR $\Rightarrow$ Quantizer

Θέμα 4 (3 μονάδες)

4.1 Εάν τα δεδομένα εισόδου σε ένα επίπεδο συνελκτικού νευρωνικού δικτύου έχουν διάσταση $200 \times 200 \times 128$ και εφαρμόσουμε φίλτρα διαστάσεων $7 \times 7 \times 128$, ποια είναι η διάσταση της εξόδου του επιπέδου αυτού, αν αποτελείται από 10 activation maps;

$$\text{input size: } H_{in} \times W_{in} \times D_{in} = 200 \times 200 \times 128 \quad H_{out} = \left\lfloor \frac{H_{in} - F_H + 2P}{S} \right\rfloor + 1 = \left\lfloor \frac{200 - 7 + 2 \cdot 0}{1} \right\rfloor + 1 = 194$$

$$\text{kernel size: } F_H \times F_W \times D_{in} = 7 \times 7 \times 128$$

$$\# \text{ kernels: } \# \text{ activation maps: } K = 10$$

$$W_{out} = \left\lfloor \frac{W_{in} - F_W + 2P}{S} \right\rfloor + 1 = \left\lfloor \frac{200 - 7 + 2 \cdot 0}{1} \right\rfloor + 1 = 194$$

$$\text{stride: } S = 1$$

$$\text{padding: } P = 0$$

$$D_{out} = K = 10$$

$$\Rightarrow \text{output size: } 194 \times 194 \times 10$$

4.2.a Τα CNNs μπορούν να μάθουν να αναγνωρίζουν ένα αντικείμενο σε μια εικόνα ανεξάρτητα από το αν το αντικείμενο μετατοπίζεται οριζόντια ή/και κάθετα.

Σωστό, είναι μερικώς μετατοπιστικά αναλλοίωτα λόγω της δομής τους.

▷ κοινά φίλτρα (shared weights)

▷ pooling layers που κάνουν αναπαράσταση ανθεκτική σε μικρές μετατοπίσεις

Ένα feature μπορεί να ανιχνεύσει και σε διαφορετικό σημείο της εικόνας

4.2.β Τα CNNs μπορούν να μάθουν να αναγνωρίζουν ένα αντικείμενο σε μια εικόνα ανεξάρτητα από το αν το αντικείμενο περιστρέφεται (στο επίπεδο εικόνας).

Λάθος, > τα CNN δεν είναι περιφερειακά αναλλοίωτα

▷ ένα φίλτρο εστιάζει μόνο σε συγκεκριμένο προσανατολισμό
μπορεί να αντιμετωπισθεί: data augmentation ή ειδικά δίκτυα

4.3 Σχεδιάζετε ένα σύστημα βαθιάς μάθησης για τη διάγνωση του καρκίνου του στήθους με ανάλυση ακτινογραφιών. Ποια πιστεύετε ότι μπορεί να είναι η καταλληλότερη μέτρηση αξιολόγησης: Accuracy, Precision, Recall; Και γιατί;

Μέτρο	Περιγραφή
Accuracy	Ποσοστό σωστών προβλέψεων επί του συνόλου (TP + TN / All).
Precision	Ποσοστό σωστών θετικών προβλέψεων επί των προβλεφθέντων θετικών (TP / TP + FP).
Recall	Ποσοστό σωστών θετικών επί των πραγματικών θετικών (TP / TP + FN).

~ Recall

Σημαντικό να εντοπιστούν όλα τα πραγματικά περιστατικά, ακόμα και αν υπάρξουν λίγες ψευδώς θετικές προβλέψεις, διότι τα ψευδώς αρνητικά είναι πιο επικίνδυνα.

4.4 Εργάζεστε ως μηχανικός, προσπαθώντας να σχεδιάσετε ένα μοντέλο που προβλέπει αν ένας χρήστης θα κάνει κλικ σε μια δεδομένη διαφήμιση. Δοκιμάζετε 2 προσεγγίσεις:

- Εκπαιδεύετε ένα μεγάλο μοντέλο, το οποίο χρειάζεται πολλές ώρες για να εκπαιδευτεί.
- Εκπαιδεύετε 10 λίγο μικρότερα μοντέλα και υπολογίζετε τον μέσο όρο των αποτελεσμάτων τους.

Και τα δύο μοντέλα σας δίνουν ακρίβεια 90% στο σετ ελέγχου. Πρέπει να διαλέξετε μία από τις δύο προσεγγίσεις για χρήση σε παραγωγική διαδικασία μέσω μιας μηχανής inference με όριο χρόνου σε πραγματικό χρόνο (real-time prediction). Ποια προσέγγιση είναι πιο κατάλληλη και γιατί;

Θα επέλεγα το ΜΕΓΑΛΟ μοντέλο.

1. ταχύτητα interface:

- ▷ μεγάλο: 1 forward pass
 - ▷ μικρό: 10 forward passes + υπολογισμός μο
- } μικρό: τουλάχιστο 10 μεγαλύτερο latency

2. υπολογιστικό κόστος:

- ▷ μεγάλο: μεγάλη μνήμη, αλλά σταθερό και προβλέψιμο κόστος
- ▷ ensemble: πολλαπλό κόστος σε CPU/GPU

3. ποιότητα αποφασής: ίδιο accuracy (90%) άρα αδιάφορο