

Πολυτεχνική Σχολή

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

**ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΨΗΦΙΑΚΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ**

**ΑΝΑΦΟΡΑ ΒΑΣΙΣΜΕΝΗ ΣΤΗΝ**

**2Η ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ**

ΑΓΓΕΛΟΣ ΝΙΚΟΛΑΟΣ ΠΟΤΑΜΙΑΝΟΣ

Α.Μ. 1084537

up1084537@ac.upatras.gr

Πάτρα, 2025

**Μοντελοποίηση Χαµηλής Τάξης**

|  |
| --- |
| **1. Αναπτύξετε στο περιβάλλον Matlab (και Python(bonus)), τον αλγόριθµο DL ακριβώς όπως αυτός ορίστηκε παραπάνω. Φροντίστε, για την δική σας ευκολία, να αναπτύξετε ξεχωριστές συναρτήσεις για την υλοποίηση του GenOMP και της KSV D.**  **2. Παρουσιάστε σε γράφημα : το συνολικό µέσο τετραγωνικό σφάλμα Y − DX2 F για κάθε εποχή, εκπαιδεύοντας το λεξικό στις training εικόνες.**  **3. (Bonus) Παρουσιάστε την αρχική και τελική µορφή του λεξικού D σε µορφή εικόνας. Δηλαδή, κάθε διάνυσµά του ϑα μετασχηματίζεται σε ένα τετραγωνικό µητρώο. Κάθε τέτοιο µητρώο ϑα εισάγεται σε ένα µεγαλύτερο µητρώο ώστε να προκύπτει µια οπτικοποιηµένη µορφή των ατόμων του λεξικού. 4. Παρουσιάστε το σφάλμα ανακατασκευής για όλες τις testing εικόνες, αφού έχετε τελειώσει µε την εκπαίδευση του λεξικού. Σχολιάστε το σφάλμα ως προς το είδος της κάθε testing εικόνας. Εάν θέλετε να χρησιμοποιήσετε δικές σας εικόνες κάντε το. Φροντίστε, ωστόσο, να διατηρείται το µοτίβο των εικόνων που σας δίνονται. Οι training εικόνες θα πρέπει να είναι ομοειδής. Ενώ οι testing θα πρέπει αν περιέχουν µία εικόνα από το testing και µία εικόνα ίδιας κλάσης µε το training.** |

Τρέξαμε τον αλγόριθμο για τις εξής παραμέτρους:

* epochs=20 - T0=5 - K=128
* epochs=20 - T0=8 - K=128
* epochs=100 - T0=5 - K=512
* epochs=200 - T0=5 - K=128
* epochs=200 - T0=8 - K=128

Το script έτρεξε για παραπάνω από συνολικά 20ώρες σε επεξεργαστή intel core i5 7600k, με την χρήση parallel processing, για την παραγωγή όλων τον αποτελεσμάτων. Τα αποτελέσματα που σχολιάζονται παρακάτω είναι για **epochs=200 - T0=8 - K=128** αλλά παρατίθενται αποτελέσματα από όλες τις παραμετροποιήσεις. Τα αποτελέσματα καθώς και τα αρχεία .fig, έτσι ώστε να φαίνονται τα patches, μπορούν να βρεθούν στον αντίστοιχο φάκελο της εργασίας με όνομα

(MATLAB) plots epochs=xxx - T0=x - K=xxx

Η παρούσα αναφορά υλοποιεί και αξιολογεί των αλγορίθμων KSVD και GenOMP στο πλαίσιο της εκμάθησης λεξικού (Dictionary Learning) και της αραιής κωδικοποίησης (Sparse Coding). Συγκεκριμένα, εφαρμόζουμε τις τεχνικές αυτές για τη μοντελοποίηση χαμηλής τάξης δεδομένων εικόνων, με στόχο την αποδοτική αναπαράσταση πληροφορίας με ελάχιστο αριθμό συνιστωσών.

**Θεωρητικό Υπόβαθρο**

Η διαδικασία της εκμάθησης λεξικού αποσκοπεί στη δημιουργία ενός συνόλου στοιχειωδών δομικών στοιχείων (ατόμων) μέσω των οποίων ένα σήμα μπορεί να ανακατασκευαστεί γραμμικά. Η εκπαίδευση ενός λεξικού D υλοποιείται μέσω της βελτιστοποίησης του προβλήματος:

όπου Y το μητρώο δεδομένων (εικόνες), DD το λεξικό και X η αραιή αναπαράσταση (Αharon et al., 2006). Το συγκεκριμένο πρόβλημα βελτιστοποίησης επιλύεται μέσω της διαδοχικής επίλυσης δύο επιμέρους προβλημάτων:

* **Αραιή Κωδικοποίηση**: Δεδομένου του λεξικού D, εύρεση του X που ικανοποιεί την αραιότητα της αναπαράστασης, χρησιμοποιώντας τον αλγόριθμο Generalized Orthogonal Matching Pursuit (GenOMP) (Wang et al., 2012).
* **Εκμάθηση Λεξικού**: Δεδομένου του αραιού κώδικα X, βελτιστοποίηση του λεξικού D μέσω του αλγορίθμου K-Singular Value Decomposition (KSVD) (Αharon et al., 2006).

Οι δύο αυτοί αλγόριθμοι επαναλαμβάνονται διαδοχικά, βελτιστοποιώντας σταδιακά το λεξικό και τους συντελεστές αναπαράστασης, επιδιώκοντας τη μείωση του σφάλματος ανακατασκευής.

**Υλοποίηση των Αλγορίθμων**

Η υλοποίηση των αλγορίθμων έγινε στο περιβάλλον MATLAB και Python. Για την αποφυγή πλεονασμών, οι υλοποιήσεις χωρίστηκαν σε ξεχωριστές συναρτήσεις για την αραιή κωδικοποίηση (GenOMP) και την εκμάθηση λεξικού (KSVD), όπως προτάθηκε από την περιγραφή της άσκησης.

**Αλγόριθμος GenOMP (Generalized Orthogonal Matching Pursuit)**

Ο αλγόριθμος GenOMP αποτελεί γενίκευση του γνωστού Orthogonal Matching Pursuit (OMP), με δυνατότητα επιλογής πολλαπλών ατόμων ανά επανάληψη, μειώνοντας τις συνολικές επαναλήψεις και τον χρόνο υπολογισμού. Η μαθηματική βάση του αλγορίθμου περιγράφεται από τους Wang et al. (2012):

* Σε κάθε επανάληψη επιλέγονται τα άτομα με τη μέγιστη συσχέτιση με το υπόλοιπο σφάλματος.
* Ακολουθεί μια ορθογώνια προβολή για την ενημέρωση του υπολοίπου (residual) σφάλματος.
* **Αλγόριθμος KSVD (K-Singular Value Decomposition)**

Ο αλγόριθμος KSVD ανανεώνει τα άτομα του λεξικού ένα προς ένα, μέσω Ανάλυσης Ιδιαζουσών Τιμών (SVD). Μετατρέπει τη διαδικασία βελτιστοποίησης σε επίλυση μικρών προβλημάτων τύπου SVD για τη βελτίωση κάθε ατόμου του λεξικού ξεχωριστά, κρατώντας σταθερό το υπόλοιπο του λεξικού (Aharon et al., 2006).

**Πειραματική Διαδικασία και Παράμετροι**

Η εκπαίδευση πραγματοποιήθηκε με τις εξής παραμέτρους:

* Μέγεθος blocks:
* Μέγιστη αραιότητα
* Αριθμός ατόμων λεξικού
* Αριθμός επαναλήψεων

Η εξέλιξη του μέσου τετραγωνικού σφάλματος (MSE) κατά την εκπαίδευση του λεξικού καταγράφηκε σε κάθε εποχή, ώστε να αξιολογηθεί η σύγκλιση του αλγορίθμου.

|  |  |
| --- | --- |
| **epochs=20 - T0=5 - K=128** | **epochs=100 - T0=5 - K=512** |
| **A graph with a line going up  AI-generated content may be incorrect.** | **A graph with a line going up  AI-generated content may be incorrect.** |

|  |  |
| --- | --- |
| **epochs=200 - T0=5 - K=128** | **epochs=200 - T0=8 - K=128** |
| **A graph of a graph  AI-generated content may be incorrect.** | **A graph with a line going up  AI-generated content may be incorrect.** |

Η καμπύλη του MSE, όπως φαίνεται στην εικόνα, παρουσιάζει μια χαρακτηριστική σύγκλιση:

* **Αρχικό στάδιο (Epochs 1-30):** Η σύγκλιση είναι ιδιαίτερα γρήγορη, από MSE=0.2704 σε MSE≈0.0336. Αυτό δείχνει ότι ο αλγόριθμος βρίσκει γρήγορα βασικά άτομα που αντιπροσωπεύουν τις κυρίαρχες δομές των δεδομένων εκπαίδευσης.
* **Μεσαίο στάδιο (Epochs 30-100):** Η ταχύτητα της σύγκλισης μειώνεται, κάτι που είναι φυσιολογικό και αναμενόμενο καθώς τα άτομα του λεξικού προσαρμόζονται σε πιο λεπτομερείς δομές.
* **Τελικό στάδιο (Epochs 100-200):** Η σύγκλιση επιβραδύνεται σημαντικά και το σφάλμα σταθεροποιείται περίπου σε 0.0256, ένδειξη ότι ο αλγόριθμος έχει φτάσει σε ένα σημείο βέλτιστου συμβιβασμού, και πιθανότατα σε κάποιο τοπικό ελάχιστο.

Αυτό το αποτέλεσμα είναι απόλυτα σύμφωνο με τη βιβλιογραφία (Aharon et al., 2006).

Found 15 training and 7 test images.  
Extracting 8x8 blocks from training set...  
Epoch 1/200 – MSE: 2.7039e-01  
Epoch 2/200 – MSE: 1.0593e-01  
Epoch 3/200 – MSE: 8.7781e-02  
Epoch 4/200 – MSE: 6.5181e-02  
Epoch 5/200 – MSE: 6.9615e-02  
Epoch 6/200 – MSE: 5.6766e-02  
Epoch 7/200 – MSE: 5.0973e-02  
Epoch 8/200 – MSE: 4.8948e-02  
Epoch 9/200 – MSE: 4.6451e-02  
Epoch 10/200 – MSE: 4.4977e-02  
Epoch 11/200 – MSE: 4.3551e-02  
Epoch 12/200 – MSE: 4.1728e-02  
Epoch 13/200 – MSE: 4.0550e-02  
Epoch 14/200 – MSE: 3.9577e-02  
Epoch 15/200 – MSE: 3.8660e-02  
Epoch 16/200 – MSE: 3.7932e-02  
…

Epoch 124/200 – MSE: 2.7352e-02  
Epoch 125/200 – MSE: 2.7401e-02  
Epoch 126/200 – MSE: 2.7287e-02  
Epoch 127/200 – MSE: 2.7268e-02  
Epoch 128/200 – MSE: 2.7243e-02  
Epoch 129/200 – MSE: 2.7228e-02  
…

Epoch 198/200 – MSE: 2.5668e-02  
Epoch 199/200 – MSE: 2.5653e-02  
Epoch 200/200 – MSE: 2.5637e-02

**Τελικό Λεξικό (Dictionary Atoms 8×8)**

Η εικόνα των ατόμων του λεξικού παρουσιάζει πολύ χαρακτηριστικά στοιχεία:

* Τα άτομα έχουν διαμορφωθεί ώστε να αναπαριστούν τυπικά δομικά μοτίβα εικόνων, όπως ακμές (edges), γραμμές (lines), καμπύλες, υφές (textures) και gradient patterns.
* Το γεγονός ότι κάθε άτομο εμφανίζει μια σαφή δομή αποτελεί ισχυρή ένδειξη επιτυχούς εκπαίδευσης.
* Είναι εμφανής η δημιουργία ενός υπερπλήρους λεξικού (overcomplete), καθώς έχουμε περισσότερα άτομα (128) από τη διάσταση των patch (64), γεγονός που εξασφαλίζει καλύτερη αραιότητα στην αναπαράσταση.

Αυτό συμφωνεί πλήρως με τις αναμενόμενες θεωρητικές προδιαγραφές που τίθενται στη βιβλιογραφία (Aharon et al., 2006).

|  |  |
| --- | --- |
| **epochs=20 - T0=5 - K=128** | **epochs=100 - T0=5 - K=512** |
|  |  |
| **epochs=200 - T0=5 - K=128** | **epochs=200 - T0=8 - K=128** |
|  |  |

**Αποτελέσματα Ανακατασκευής (Reconstruction)**

|  |
| --- |
| **epochs=200 - T0=8 - K=128** |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

Η διαφορά και το dictionary γίνονται ακόμα πιο αισθητά όταν κάνουμε zoom in στα αποτελέσματα:

|  |
| --- |
| **epochs=20 - T0=5 - K=128** |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **epochs=200 - T0=5 - K=128** |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **epochs=100 - T0=5 - K=512** |
|  |
|  |
|  |
|  |
|  |

**Εικόνα: Γάτα (γνωστή εικόνα εκπαίδευσης)**

* **PSNR = 34.72 dB**
* Η ανακατασκευή είναι πολύ υψηλής ποιότητας. Οι λεπτομέρειες στα μάτια και στο τρίχωμα αναπαράγονται με ακρίβεια.
* Ο θόρυβος και τα artifacts είναι ελάχιστα έως ανύπαρκτα, ένδειξη ότι το λεξικό έχει μάθει καλά τη συγκεκριμένη κλάση (γατών).

**Εικόνα: Γάτα (άγνωστη εικόνα)**

* **PSNR = 30.69 dB**
* Πολύ καλή αναπαραγωγή, αλλά εμφανής η μείωση στην ποιότητα σε σχέση με την προηγούμενη. Ελαφρά θόλωση (blurring) και απώλεια λεπτομέρειας γύρω από τα μάτια και το τρίχωμα.
* Αναμενόμενο αποτέλεσμα, καθώς η εικόνα ανήκει στην ίδια γενική κατηγορία με τις training εικόνες, αλλά το λεξικό είναι λιγότερο προσαρμοσμένο στις λεπτομέρειες της συγκεκριμένης εικόνας.

**Εικόνα: Field**

* **PSNR = 24.85 dB**
* Η ανακατασκευή έχει εμφανείς δυσκολίες στις λεπτομέρειες υψηλής συχνότητας, όπως τα φύλλα των δέντρων και το γρασίδι.
* Αισθητή υποβάθμιση των λεπτομερειών και παρουσία artifacts που υποδεικνύουν ότι το λεξικό δεν είναι προσαρμοσμένο για τη συγκεκριμένη κατηγορία εικόνας. Είναι λογικό, αφού το λεξικό εκπαιδεύτηκε κυρίως σε εικόνες με γάτες.

**Εικόνα: Πορτρέτο Γυναίκας**

* **PSNR = 36.81 dB**
* Εξαιρετικά υψηλή ποιότητα ανακατασκευής. Οι λεπτομέρειες στο πρόσωπο, τα μαλλιά και ιδιαίτερα η υφή του δέρματος έχουν ανακατασκευαστεί πολύ πιστά.
* Η αραιή κωδικοποίηση είναι ιδιαίτερα επιτυχημένη εδώ, κάτι που υποδεικνύει ότι το λεξικό διαθέτει επαρκείς γενικές δομές (άτομα) για την επιτυχή ανακατασκευή ανθρώπινων προσώπων.

**Εικόνα: Πορτρέτο Γυναίκας (λεπτομέρεια προσώπου)**

* **PSNR = 36.81 dB**
* Ιδιαίτερα επιτυχημένη ανακατασκευή ακόμα και σε επίπεδο λεπτομέρειας. Η υφή του δέρματος, οι φακίδες και οι λεπτομέρειες του μακιγιάζ είναι πολύ κοντά στην πραγματική εικόνα.
* Αποδεικνύεται η αποτελεσματικότητα του λεξικού στη λεπτομερή και υψηλής ποιότητας αναπαραγωγή λεπτών χαρακτηριστικών.

**Συμπεράσματα και Παρατηρήσεις**

* Το λεξικό που εκπαιδεύτηκε είναι εξαιρετικά αποδοτικό για την αναπαράσταση εικόνων παρόμοιων με αυτές της εκπαίδευσης (π.χ., γάτες), αλλά παρουσιάζει δυσκολίες σε εικόνες με σημαντικά διαφορετικά χαρακτηριστικά (τοπία).
* Η επιλογή του T0 = 8 προσφέρει έναν καλό συμβιβασμό ανάμεσα στην αραιότητα και την ποιότητα αναπαράστασης.
* Υψηλά PSNR σημαίνουν ότι οι εικόνες ανακατασκευάζονται με ελάχιστο σφάλμα, κάτι που συμβαδίζει απόλυτα με τη θεωρία της αραιής κωδικοποίησης (Wang et al., 2012).
* Παρατηρείται ότι η επιτυχία της ανακατασκευής εξαρτάται έντονα από το είδος των εικόνων που χρησιμοποιήθηκαν για την εκπαίδευση του λεξικού, γεγονός που επιβεβαιώνει τη σημασία της επιλογής ομοιογενών δεδομένων εκπαίδευσης.

**Τελική εκτίμηση της μεθοδολογίας**

Η εφαρμογή των αλγορίθμων KSVD και GenOMP αποδείχθηκε εξαιρετικά αποτελεσματική για την αραιή αναπαράσταση και ανακατασκευή εικόνων. Επισημαίνεται, ωστόσο, η ανάγκη για κατάλληλη επιλογή δεδομένων εκπαίδευσης, που να είναι αντιπροσωπευτικά των εικόνων που επιθυμούμε να αναπαραστήσουμε. Για περαιτέρω βελτίωση:

* Εκπαίδευση του λεξικού σε μεγαλύτερο και πιο ποικίλο dataset.