

Ε.Μ.Π. - Σχολή Ηλεκτρολόγων Μηχ/κών & Μηχ/κών Υπολογιστών

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ¹ 2024-2025

Καθηγητής Πέτρος Μαραγκός, E-mail: maragos@cs.ntua.gr
Εργαστήριο: CVSP: <http://cvsp.cs.ntua.gr>, IRAL: <http://robotics.ntua.gr>

Οι ενδιαφερόμενοι/ες παρακαλούνται να αποστείλουν ηλεκτρονικά σε ένα Zip με το όνομά τους τα εξής:
i) την αναλυτική τους βαθμολογία από ΣΗΜΜΥ-ΕΜΠ με μια εκτίμηση του τρέχοντος μέσου όρου, ii) το βιογραφικό τους, και iii) τον αριθμό των οφειλομένων μαθημάτων αν βρίσκονται στο 4^ο ή ανώτερο έτος σπουδών. Η αποστολή της πληροφορίας αυτής να γίνεται στην Γραμματέα του Εργαστηρίου κα. Βίκυ Πλατίτσα (email: <vickyplatitsa@gmail.com>).

Προαπαιτούμενα μαθήματα (Ροής Σ) με επιτυχή και συνεπή παρακολούθηση και πολύ καλή απόδοση:

- Για θέματα N: ΨΕΣ, και Μηχανική Μάθηση-MM ή Αναγνώριση Προτύπων-ΑΠ (*).
- Για θέματα V: Οραση Υπολογιστών (ΟΥ), ή πιθανώς και MM ή ΑΠ (*).
- Για θέματα R που συνδυάζουν Ρομποτική με CV/SP/ML: ΟΥ και Ρομποτική I ή MM/ΑΠ (*).
- Για θέματα AVM: ΨΕΣ, ΟΥ και MM/ΑΠ (*).
- Για θέματα L: ΨΕΣ, ΟΥ και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα S: ΨΕΣ και κάποιο μάθημα ΣΑΕ απο Ροη Σ, ή πιθανώς και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα T: ΟΥ και MM/ΑΠ (*).

* = μπορεί να παρακολουθείται ταυτόχρονα με διπλωματική.

- Γενικά, για να υπάρχει ποικιλία ερευνητικών θεμάτων, δίνονται περισσότερα θέματα από τις θέσεις που μπορούν να επιβλεφθούν εντός ενός έτους. **Τα ενεργά θέματα εντός του 2024-2025 είναι αυτά με το σήμα (*) στον τίτλο τους.**
- Μετά την εκδήλωση ενδιαφέροντος οι ενδιαφερόμενοι φοιτητές θα συναντηθούν μαζί με τον υπεύθυνο καθ. Π. Μαραγκό και Επιστημονικούς/ές Συνεργάτες/ιδες καθώς και Υποψήφιους/ες Διδάκτορες της ερευνητικής ομάδας και θα γίνει προσπάθεια να αντιστοιχισθούν θέματα και φοιτητές με όσο το δυνατόν καλύτερο ταίριασμα ενδιαφερόντων και ικανοτήτων.
- Η τελική αποδοχή αίτησης για εκπόνηση διπλωματικής θα εξαρτηθεί από την επίδοση στα σχετικά μαθήματα του εργαστηρίου, τον αριθμό των οφειλομένων μαθημάτων, και την χρονική διαθεσιμότητα του θέματος.

Physiological Data Processing and Learning from Wearable Sensors for Healthcare Applications:

Η τεχνολογική πρόοδος που έχει σημειωθεί τα τελευταία χρόνια στις φορητές συσκευές (κινητά τηλέφωνα, smart-watch, κτλ.) έχει πλέον επιτρέψει την αξιόπιστη καταγραφή πολλών και διαφορετικών βιομετρικών δεικτών, καθώς επίσης και ακουστικών σημάτων, τα οποία πλέον χρησιμοποιούνται και στην ιατρική.

(*) N1. Ανάλυση χρονοσειρών βιομετρικών δεικτών για εντοπισμό και πρόβλεψη υποτροπών σε ασθενών με ψυχωτικές διαταραχές

Στόχος της παρούσας διπλωματικής είναι η ανάπτυξη έξυπνων αλγορίθμων που θα εντοπίζουν και θα προβλέπουν υποτροπές σε ασθενείς με ψυχωτικές διαταραχές μέσω της αξιοποίησης δεδομένων που καταγράφονται από έξυπνες-φορητές συσκευές κατά τη διάρκεια της καθημερινότητας των ασθενών. Η αξιοποίηση της μηχανικής μάθησης σε βιομετρικά δεδομένα μπορεί να δώσει το συμπεριφορικό προφίλ των ασθενών (digital phenotyping) και στη συνέχεια τον εντοπισμό αποκλίσεων συμπεριφορών από το σύνηθες μοτίβο του κάθε ασθενούς [2,3], μεταβολών και τάσεων που θα αποκαλύψουν μια ψυχωτική υποτροπή. Η σπανιότητα των υποτροπών και η έλλειψη μεγάλων επισημειωμένων συνόλων δεδομένων περιορίζουν την αποτελεσματικότητα της επιβλεπόμενης μάθησης, αναδεικνύοντας την ανάγκη για καινοτόμες προσεγγίσεις όπως η ανίχνευση ανωμαλιών [1,4]. Η πρόβλεψη μιας τέτοιας υποτροπής μπορεί να αποτελέσει ένα σημαντικό εργαλείο στην έγκαιρη αντιμετώπιση της.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου, Δρ. Παναγιώτης Φιλντίσης ΕΜΠ και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

¹ Μερικά από τα ανωτέρω θέματα έχουν προοπτική για Διδακτορικό με οικονομική υποστήριξη από ερευνητικά προγράμματα.

References:

- [1] N. Efthymiou, G. Retsinas, P. Filntisis, and P. Maragos, [Augmenting Transformer Autoencoders with Phenotype Classification for Robust Detection of Psychotic Relapses](#), in Proc. IEEE International Conference on Health Informatics (ICASSP), 2024.
- [2] N. Efthymiou, G. Retsinas, P. Filntisis, C. Garoufis, A. Zlatintsi, E. Kalisperakis, V. Garyfalli, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, [From digital phenotype identification to detection of psychotic relapses](#), in Proc. IEEE International Conference on Health Informatics (ICHI), 2023.
- [3] E. Fekas, A. Zlatintsi, P. Filntisis, C. Garoufis, N. Efthymiou, and P. Maragos, [Relapse prediction from long-term wearable data using self-supervised learning and survival analysis](#), in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023
- [4] G. Retsinas, P.P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi and P. Maragos, [Person Identification Using Deep Convolutional Neural Networks On Short-term Signals From Wearable Sensors](#), in Proc. 45th IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020), May 2020.
- [5] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, “[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#),” *Sensors*, vol. 22, no. 19, Oct. 2022.
- [6] C. Lee & M. Van Der Schaar, [Temporal phenotyping using deep predictive clustering of disease progression](#). In *Proc. International Conference on Machine Learning (ICML)*, 2020.
- [7] D. A. Adler et al., [Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks](#), *JMIR mHealth and uHealth*, Vol. 8, 2020.

(*) N2. Ανάλυση χρονοσειρών βιομετρικών δεικτών και βίντεο δεδομένων για τον εντοπισμό καταθλιπτικών υποτροπών

Ένα σύγχρονο πεδίο έρευνας που αναπτύσσεται και επεκτείνεται διαρκώς είναι αυτό της ηλεκτρονικής υγείας (eHealth). Η αυξανόμενη χρήση wearable και audiovisual συσκευών έχει ως αποτέλεσμα τη συλλογή μεγάλου όγκου πολυμεσικών δεδομένων που καταγράφουν πολλές πτυχές της καθημερινότητας των ανθρώπων. Ο συνδυασμός αυτών των δεδομένων με την αλματώδη ανάπτυξη της μηχανικής μάθησης έχουν ανοίξει νέες κατευθύνσεις στην ανάλυση σημάτων από πολλαπλούς αισθητήρες για την εκτίμηση της ψυχικής-συναισθηματικής κατάστασης του ατόμου. Στην κατεύθυνση αυτή, η διπλωματική στοχεύει στην ανάπτυξη αλγορίθμων που αξιοποιούν τόσο βιοσήματα όσο και οπτική πληροφορία από βίντεο με σκοπό τον εντοπισμό καταθλιπτικών υποτροπών.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου, Δρ. Παναγιώτης Φιλντίσης ΕΜΠ)

References:

- [1] N. Efthymiou, G. Retsinas, P. Filntisis, and P. Maragos, [Augmenting Transformer Autoencoders with Phenotype Classification for Robust Detection of Psychotic Relapses](#), in Proc. IEEE International Conference on Health Informatics (ICASSP), 2024.
- [2] N. Efthymiou, G. Retsinas, P. Filntisis, C. Garoufis, A. Zlatintsi, E. Kalisperakis, V. Garyfalli, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, [From digital phenotype identification to detection of psychotic relapses](#), in Proc. IEEE International Conference on Health Informatics (ICHI), 2023.
- [3] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, “[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#),” *Sensors*, vol. 22, no. 19, Oct. 2022.
- [4] Giannakakis, G., Koujan, M.R., Roussos, A. *et al.* [Automatic stress analysis from facial videos based on deep facial action units recognition](#). *Pattern Anal Applic* 25, 521–535 (2022).
- [5] Pedrelli, P., Fedor, S., Ghandeharioun, A., Howe, E., Ionescu, D.F., Bhatena, D., Fisher, L.B., Cusin, C., Nyer, M., Yeung, A. and Sangermano, L. [Monitoring changes in depression severity using wearable and mobile sensors](#). *Frontiers in psychiatry*, 11, p.584711.

(*) N3. Ανάλυση ηχητικών σημάτων για υποστήριξη ασθενών σε θέματα ψυχικής υγείας.

Αντικείμενο αυτής της διπλωματικής εργασίας είναι ο εντοπισμός, ή και η πρόβλεψη, υποτροπιαζουσών καταστάσεων σε ασθενείς στο διπολικό ή στο σχιζοειδές φάσμα χρησιμοποιώντας βασιζόμενοι σε δεδομένα ομιλίας των ασθενών κατά τη διάρκεια σύντομων κλήσεων με θεράποντες ιατρούς [1-5]. Ενδεικτικές κατευθύνσεις αποτελούν 1) η χρήση παράλληλων τροπικοτήτων σε συνδυασμό με την ομιλία, όπως για παράδειγμα καταγραφές κειμένου των συνομιλιών [6], 2) εκμετάλλευση προεκπαιδευμένων γλωσσικών/ακουστικών μοντέλων για το σκοπό αυτό [7,8], ή μελέτη της επίδρασης βάσεων δεδομένων που έχουν συλλεχθεί για συγγενικά προβλήματα, όπως η αναγνώριση συναισθηματικής κατάστασης [9, 10].

(Συνεργάτες: ΥΔ Χρήστος Γαρούφης, Δρ. Νάνσυ Ζλατίντση, ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

- [1] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#)," *Sensors*, vol. 22, no. 19, Oct. 2022.
- [2] Garoufis, C., Zlatintsi, A., Filntisis, P. P., Efthymiou, N., Kalisperakis, E., Karantinos, T., ... & Maragos, P. [Towards unsupervised subject-independent speech-based relapse detection in patients with psychosis using variational autoencoders](#). In *Proc. EUSIPCO 2022*
- [3] J. Gideon, K. Matton, S. Anderau, M.G. McInnis, and E.M. Provost, "[When to Intervene: Detecting Abnormal Mood using Everyday Smart-phone Conversations](#)," arXiv preprint arXiv:1909.11248, 2019.
- [4] K.-Y. Huang, C.-H. Wu, and M.-H. Su, "[Attention-Based Convolutional Neural Network and Long Short-Term Memory for Short-Term Detection of Mood Disorders Based on Elicited Speech Responses](#)," *Pattern Recognition*, vol. 88, pp. 668–678, 2019.
- [5] Liu, S., A. Mallol-Ragolta, and B. W. Schuller. "[COVID-19 Detection from Speech in Noisy Conditions](#)." In *Proc. ICASSP 2023*.
- [6] Yoon, S., Dey, S., Lee, H., & Jung, K. [Attentive modality hopping mechanism for speech emotion recognition](#). In *Proc. ICASSP 2020*
- [7] Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., ... & Wei, F. [Wavlm: Large-scale self-supervised pre-training for full stack speech processing](#). *IEEE Journal of Selected Topics in Signal Processing*, 16(6), 1505-1518.
- [8] Morais, E., Hoory, R., Zhu, W., Gat, I., Damasceno, M., & Aronowitz, H. [Speech emotion recognition using self-supervised features](#). In *Proc. ICASSP 2022*
- [9] Ringeval, F. et al., "[AVEC 2017: Real-life depression, and affect recognition workshop and challenge](#)." In *Proceedings of the 7th annual workshop on audio/visual emotion challenge*
- [10] Wang, K. et al., "[MEAD: A large-scale audio-visual dataset for emotional talking-face generation](#)." In *Proc. ECCV 2020*

Επικοινωνία Ανθρώπου-Ρομπότ και Αυτόνομη Πλοήγηση: Πολυτροπικές Ευφυείς Διεπαφές, Οραση, Ανάλυση Οπτικο-Ακουστικών Σημάτων, Ανίχνευση, Αναγνώριση, Πλοήγηση:

(*) R1. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ, με εφαρμογές σε assistive robotics:

Μια από τις προκλήσεις για την ανάπτυξη ευφύων ρομπότ που αλληλεπιδρούν με φυσικό τρόπο με το περιβάλλον τους, είναι η ικανότητά τους να αναγνωρίζουν τις δράσεις και τις χειρονομίες που εκτελούν οι άνθρωποι ανά πάσα στιγμή. Στόχος της διπλωματικής αυτής είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών σε πραγματικό χρόνο, η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων, όπως RGB-D ή/και ανθρώπινη πόζα. Οι μέθοδοι αυτές μπορούν να χρησιμοποιηθούν σε μεγάλες βάσεις δεδομένων ή/και σε ρομποτικές διατάξεις υποβοήθησης ηλικιωμένων και ατόμων με κινητικά προβλήματα.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης)

References:

- [1] I. Rodomagoulakis, N. Kardaris, V. Pitsikalis, E. Mavroudi, A. Katsamanis, A. Tsiami and P. Maragos, "[Multimodal Human Action Recognition in Assistive Human-Robot Interaction](#)", Proc. ICASSP-2016, Shanghai, China, Mar. 2016.
- [2] A. Zlatintsi, I. Rodomagoulakis, V. Pitsikalis, P. Koutras, N. Kardaris, X. Papageorgiou, C. Tzafestas and P. Maragos, "[Social Human-Robot Interaction for the Elderly: Two Real-life Use Cases](#)", Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI-2017), Vienna, Austria, March 2017.
- [4] J. Carreira, & A. Zisserman, "[Quo vadis, action recognition? a new model and the kinetics dataset](#)". Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017), Honolulu, Hawaii, July 2017
- [5] M. Xu, M. Gao, Y. Chen, L. S. Davis, D. J. Crandall, "[Temporal recurrent networks for online action detection](#)", Proc. IEEE Int'l Conference on Computer Vision (ICCV-2019), Seoul, Korea, Oct.-Nov. 2019.
- [6] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, Ahmed A. A. Osman, D. Tzionas, and M. J. Black, "[Expressive Body Capture: 3D Hands, Face, and Body from a Single Image](#)", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2019), Long Beach, CA, United States, June 2019.
- [7] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "[NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding](#)", in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2019
- [8] G. Chalvatzaki, P. Koutras, A. Tsiami, C. S. Tzafestas and P. Maragos, "[i-Walk Intelligent Assessment System: Activity, Mobility, Intention, Communication](#)", Proc. 16th European Computer Vision Conference Workshops (ECCVW) – 8th Int'l Workshop on Assistive Computer Vision and Robotics (ACVR-2020), Aug. 2020.
- [9] V. Vasileiou, N. Kardaris, P. Maragos, "[Exploring Temporal Context and Human Movement Dynamics for Online Action Detection in Videos](#)", in Proc. 29th European Signal Processing Conference (EUSIPCO 2021), Dublin, Ireland, 2021.
- [10] G. Moustiris, N. Kardaris, A. Tsiami, G. Chalvatzaki, P. Koutras, A. Dometios, P. Oikonomou, C. Tzafestas, P. Maragos, E. Efthimiou, X. Papageorgiou, S.-E. Fotinea, Y. Koumpouros, A. Vacalopoulou, E. Papageorgiou, A. Karavasili, F. Koureta, D. Dimou, A. Nikolakakis, K. Karaiskos and P. Mavridis, "[The i-Walk Lightweight Assistive Rollator: First Evaluation Study](#)", in *Frontiers in Robotics and AI*, vol. 8, pp. 272, Sep. 2021.

(* R2. Αναγνώριση ανθρώπινων δράσεων κατά την αλληλεπίδραση ανθρώπων ή ανθρώπου-ρομπότ σε πραγματικό χρόνο με εφαρμογή στην ανίχνευση νευροαναπτυξιακών διαταραχών σε παιδιά

Στόχος είναι η ανάπτυξη ενός συστήματος που θα εκτιμά αυτόματα το μοτίβο αλληλεπίδρασης και εμπλοκής μεταξύ των δύο συμμετεχόντων (παιδιού-μητέρας ή παιδιού-ρομπότ)[1,2]. Το σύστημα αυτό θα βασίζεται κύρια σε δεδομένα βίντεο [3,4] από τα οποία μπορεί να εξάγονται χαρακτηριστικά για την πόζα, τις χειρονομίες, τις εκφράσεις του προσώπου, και το βλέμμα αλλά και επιπρόσθετα πολυτροπικά δεδομένα σχετικά με την κίνηση και την λεκτική συμπεριφορά. Συνολικά θα γίνει εκτίμηση της συνεισφοράς των διαφόρων τροποικοτήτων στην εκτίμηση της εμπλοκής των παιδιών [5]. Τα μοτίβα διάδρασης των παιδιών θα αξιοποιηθούν τελικώς στην εκμάθηση του μοντέλου για την ανίχνευση εμφάνιση νευροαναπτυξιακών διαταραχών.

(Συνεργάτες: Δρ. Νίκη Ευθυμίου)

References:

- [1] D. Anagnostopoulou, N. Efthymiou, C. Papailiou and P. Maragos, "[Child Engagement Estimation in Heterogeneous Child-Robot Interactions Using Spatiotemporal Visual Cues](#)", Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2022), Kyoto, Japan, October 2022.
- [2] Rudovic, O., Lee, J., Dai, M., Schuller, B., & Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3, eaao6760.
- [3] N. Efthymiou, P. P. Filintisis, G. Potamianos, and P. Maragos, "[Visual Robotic Perception System with Incremental Learning for Child-Robot Interaction Scenarios](#)", Technologies, vol. 9, no. 4, article 86, Nov. 2021.
- [4] E. Marinoiu, M. Zanzfir, V. Olaru, and C. Sminchisescu. "[3D human sensing, action and emotion recognition in robot assisted therapy of children with autism](#)", Proc. CVPR 2018.
- [5] Akbari, H., Yuan, L., Qian, R., Chuang, W. H., Chang, S. F., Cui, Y., & Gong, B. (2021). Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text. *Advances in Neural Information Processing Systems*, 34, 24206-24221.

(* R3. Visual SLAM systems with geometric and/or semantic enhancements:

Τα συστήματα vSLAM [1,2] επιτυγχάνουν ταυτόχρονα την 3D ανακατασκευή του άγνωστου χώρου στον οποίο κινούνται, καθώς και την εκτίμηση της τροχιάς που διαγράφουν. Το vSLAM αποτελεί τη real-time εκδοχή του κλασικού προβλήματος της ΟΥ, Structure from Motion (SfM), ενώ έχει επίσης μελετηθεί παλιότερα υπό τη μορφή Time-of-Flight (ToF) μεθόδων (lasers ή sonars) και στο άλλο κλασικό πρόβλημα της Ρομποτικής, το SLAM. Πλέον, τα παραπάνω ενοποιούνται σε ένα σύστημα, με χρήση mono, stereo ή RGBD καμερών, καθώς

και είναι δυνατό να συνδυαστούν με πληροφορία από LIDARs ή IMUs [7] οδηγώντας σε παραλλαγές συστημάτων vSLAM.

Έμφαση δίνεται:

- στην αξιοποίηση συνθετότερων γεωμετρικών δομών της εικόνας (ευθείες, επίπεδα, κωνικές) [3,4,6],
- στην ενσωμάτωση δυναμικών άκαμπτων αντικειμένων με ταυτόχρονο υπολογισμό της πόζας τους στον χρόνο [8],
- στη χρήση σημείων φυγής για αποσύζευξη μεταφορικής και περιστροφικής συνιστώσας της πόζας [9], καθώς και
- στην εξαγωγή και αξιοποίηση σημασιολογικής πληροφορίας [4,5].

Τα ανωτέρω υλοποιούνται με σκοπό την αύξηση της ευρωστίας, της ακρίβειας, του εύρους των περιβαλλόντων εφαρμογής, ή τη μείωση της υπολογιστικής πολυπλοκότητας. Ευρέως χρησιμοποιούμενα σύνολα δεδομένων σε εξωτερικούς χώρους είναι το [10] και σε εσωτερικούς τα [11,12].

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας)

References:

- [1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, *IEEE Transactions on Robotics*, 2015.
- [2] D. Schlegel, M. Colosi and G. Grisetti, “ProSLAM: Graph SLAM from a Programmer's Perspective”, *IEEE International Conference on Robotics and Automation*, 2018.
- [3] F. Nardi, B. D. Corte and G. Grisetti, “Unified Representation and Registration of Heterogeneous Sets of Geometric Primitives”, *IEEE Robotics and Automation Letters*, 2019.
- [4] L. Nicholson, M. Milford and N. Sünderhauf, “QuadricSLAM: Dual Quadrics From Object Detections as Landmarks in Object-Oriented SLAM”, *IEEE Robotics and Automation Letters*, 2019.
- [5] I. Asmanis, P. Mermigkas, G. Chalvatzaki, J. Peters and P. Maragos, “A Semantic Enhancement of Unified Geometric Representations for Improving Indoor Visual SLAM”, *19th International Conference on Ubiquitous Robots (UR)*, 2022.
- [6] M. Hosseinzadeh, Y. Latif, T. Pham, N. Sünderhauf and I. Reid, “Structure Aware SLAM Using Quadrics and Planes”, *ACCV*, 2018.
- [7] A. Rosinol, M. Abate, Y. Chang and L. Carlone, “Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping”, *IEEE International Conference on Robotics and Automation*, 2020.
- [8] A. Manetas, P. Mermigkas and P. Maragos, “SDPL-SLAM: Introducing Lines in Dynamic Visual SLAM and Multi-Object Tracking”, *Proc. of IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2024.
- [9] J. Liu and Z. Meng, “Visual SLAM With Drift-Free Rotation Estimation in Manhattan World,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6512-6519, Oct. 2020.
- [10] KIT, Visual Odometry/SLAM Evaluation, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php
- [11] TUM, RGB-D SLAM Dataset and Benchmark, [Online]. Available: <https://vision.in.tum.de/data/datasets/rgbd-dataset>
- [12] ICL-NUIM, Dataset for benchmarking VO and SLAM algorithms, [Online]. Available: <https://www.doc.ic.ac.uk/~ahanda/VaFRIC/iclnuim.html>

(* R4. Visual localization with Deep Learning techniques using RGB images for robotic inspection in power substations

Στα πλαίσια της διπλωματικής επιθυμούμε την απόκτηση ενός πυκνού νέφους σημείων για την αναπαράσταση του χώρου ενός Κέντρου Υψηλής Τάσης (KYT) (π.χ. pointcloud που έχει παραχθεί με μεθόδους LiDAR SLAM ή visual SLAM) και κατόπιν την επίλυση του προβλήματος της αυτόματης χωροθέτησης μιας άγνωστης εικόνας προερχόμενης από το ίδιο KYT. Η παραπάνω μεθοδολογία που συναντάται με το όνομα visual localization βρίσκει εφαρμογές σε πεδία όπως η αυτόνομη πλοήγηση, η επαυξημένη πραγματικότητα [1], κ.ά. Έρευνες έχουν εστιάσει στη χρήση μικρού συνόλου εικόνων (few-shot learning) [2], στην ταυτόχρονη χρήση τρισδιάστατης και σημασιολογικής πληροφορίας [3], καθώς και πρόσφατα στην ενσωμάτωση 3D Gaussian Splatting [1]. Το θέμα αυτό μπορεί να εφαρμοστεί πάνω στην αυτόματη επισκόπηση ηλεκτρολογικού εξοπλισμού σε κέντρα υψηλής τάσης με τη βοήθεια αυτόνομου ρομπότ.

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας, Νίκος Κάρδαρης)

- [1] Zhai, Hongjia et al, “SplatLoc: 3D Gaussian Splatting-based Visual Localization for Augmented Reality”, 10.48550/arXiv.2409.14067.
- [2] Dong, Siyan, et al, "Visual localization via few-shot scene region classification", *Proc. International Conference on 3D Vision (3DV)*, 2022.
- [3] Schönberger, Johannes L. et al, "Semantic visual localization", *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

(*) R5. Lidar-Camera Registration and multimodal fusion using geometric and/or Deep Learning approaches

Μια σημαντική πρόκληση σε πολυαισθητηριακά ρομποτικά συστήματα είναι η αυτόματη βαθμονόμηση (registration/calibration) των αισθητήρων μεταξύ τους, δηλαδή η εύρεση του μετασχηματισμού που συνδέει τα πλαίσια/συστήματα συντεταγμένων τους. Συχνά αυτό αποτελεί προϋπόθεση για την από κοινού χρήση ή πολυτροπική ανάλυση δεδομένων που μπορεί να είναι ετερογενή, όπως IMU/Lidar/κάμερες. Το συγκεκριμένο θέμα αφορά τη βαθμονόμηση έγχρωμης κάμερας με αισθητήρα Lidar οι οποίοι είναι τοποθετημένοι σε κινούμενο ρομπότ, με χρήση μεθόδων χωρίς στόχο ("targetless"). Πολλές προσεγγίσεις εντοπίζουν σημειακές αντιστοιχίσεις (matches) μεταξύ μιας έγχρωμης εικόνας και ενός τρισδιάστατου νέφους σημείων (point cloud) [1,2], ή αντιστοιχίζουν γραμμές/επίπεδα [3,4], ενώ άλλες αντιμετωπίζουν το πρόβλημα με end-to-end τρόπο, τυπικά με χρήση βαθιών νευρωνικών δικτύων [5]. Στα πλαίσια της διπλωματικής μπορεί να δοθεί έμφαση στην εξαγωγή χαρακτηριστικών για την εύρεση αντιστοιχίσεων, στο συνδυασμό σημείων/γραμμών/επιπέδων για τη βελτιστοποίηση της εκτιμώμενης πόζας, ή την ενσωμάτωση γεωμετρικής πληροφορίας στην εκπαίδευση των βαθιών νευρωνικών δικτύων.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης, Παναγιώτης Μέρμιγκας)

- [1] Koide et al., "General, Single-shot, Target-less, and Automatic LiDAR-Camera Extrinsic Calibration Toolbox", in Proc. International Conference on Robotics and Automation, 2023.
- [2] Lv, Xudong, et al. "LCCNet: LiDAR and camera self-calibration using cost volume network", in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [3] Zhou, Lipu, Zimo Li, and Michael Kaess. "Automatic Extrinsic Calibration of a Camera and a 3D LiDAR Using Line and Plane Correspondences." In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018
- [4] Z. Luo, G. Yan, X. Cai and B. Shi, "Zero-training LiDAR-Camera Extrinsic Calibration Method Using Segment Anything Model", in Proc. IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 2024.
- [5] Iyer, Ganesh, et al, "CalibNet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks", in Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018.

(*) R6. Representation Learning of Visual Entities and Generalization using Reinforcement Learning.

Εξαγωγή οπτικών χαρακτηριστικών (π.χ. σχήμα, χρώμα) απλών αντικειμένων (π.χ. σφαίρες, κύβοι) μέσω Reinforcement Learning σε περιβάλλοντα προσομοίωσης όπου ένας ρομποτικός βραχίονας αλληλεπιδρά με τα αντικείμενα αυτά, χωρίς καμία πρότερη γνώση των αντικειμένων. Επί της ουσίας θα «χτιστεί» ένα σύστημα όρασης από δεδομένα προσομοίωσης και θα γίνει μελέτη ικανότητας γενίκευσης εξαγόμενων χαρακτηριστικών.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] Heuillet, Alexandre, Fabien Couthouis, and Natalia Díaz-Rodríguez. "Explainability in deep reinforcement learning." *Knowledge-Based Systems* 214 (2021): 106685.
- [2] Gershman, Samuel, Jonathan Cohen, and Yael Niv. "Learning to selectively attend." *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 32. No. 32. 2010.
- [3] Le Goff, Leni K., et al. "Building an affordances map with interactive perception." *Frontiers in Neurorobotics* 16 (2022): 504459.

(*) R7: Development and Experimentation with a TIAGo Robot for Grasping and Moving Items Using 3D Reconstruction and 6D Pose Estimation Methods

Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη και ο πειραματισμός ενός ρομποτικού συστήματος με τη χρήση του ρομπότ **TIAGo** της **PAL Robotics** [1], το οποίο θα μπορεί να ανιχνεύει, να πιάνει και να μετακινεί αντικείμενα που βρίσκονται πάνω σε ένα τραπέζι. Το ρομπότ θα αξιοποιήσει προηγμένες τεχνικές 3D ανακατασκευής και αλγορίθμους εκτίμησης της 6D πόζας των αντικειμένων. Οι μέθοδοι αυτές θα επιτρέψουν στο ρομπότ να αναγνωρίζει με ακρίβεια τη θέση και τον προσανατολισμό των αντικειμένων στον τρισδιάστατο χώρο, ώστε να τα μετακινεί σύμφωνα με προδιαγεγραμμένα μοτίβα κίνησης. Η διπλωματική εργασία συνδυάζει τεχνολογίες όρασης υπολογιστών, ρομποτικής, και την ανάπτυξη και ενσωμάτωση λογισμικού για τη λειτουργία του ρομπότ.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] PAL Robotics, TIAGo Robot - A Mobile Manipulator for Research and Industry. Available at: <https://pal-robotics.com/robot/tiago/>
- [2] Labbe, Y., Garcia-Gasulla, D., Carpentier, J., Aubry, M., & Laptev, I. (2022). MegaPose: 6D Pose Estimation of Novel Objects via Render & Compare. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14362-14372).
- [3] Mousavian, A., Eppner, C., Fox, D., & Cosgun, A. (2019). 6-DOF GraspNet: Variational Grasp Generation for Object Manipulation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2901-2910).

Οραση Υπολογιστών:

(*) **V1. Υλοποίηση συστήματος αναγνώρισης συμπεριφοράς (συναισθηματικής κατάστασης/χειρονομιών/δράσεων) και κατάτμηση εικόνων σε διαφορετικές σημασιολογικές περιοχές (π.χ. βλάστηση, δρόμοι, άνθρωποι) με χρήση βαθέων νευρωνικών δικτύων και μεθόδων domain adaptation και integration του συστήματος στο Robot-Operating-System software.**

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ, και Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos and P. Maragos, "[Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child-Robot Interaction](#)," IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 4011-4018, Oct. 2019. DOI: [10.1109/LRA.2019.2930434](https://doi.org/10.1109/LRA.2019.2930434) .
- [2] <https://www.ros.org/>
- [3] H. Gunes, & M. Pantic, "Automatic, dimensional and continuous emotion recognition", ACM International Journal of Synthetic Emotions, 1(1), 68-99, 2010. <https://dl.acm.org/doi/10.4018/jse.2010101605> .
- [4] P. P. Filntisis, N. Efthymiou, G. Potamianos, P. Maragos, "[Emotion Understanding in Videos Through Body, Context, and Visual-Semantic Embedding Loss](#)", Proc. ECCV Workshop, 2020.
- [5] I. Pikoulis, P. P. Filntisis, and P. Maragos. "[Leveraging Semantic Scene Characteristics and Multi-Stream Convolutional Architectures in a Contextual Approach for Video-Based Visual Emotion Recognition in the Wild](#)," Proc. FG Conference, 2021
- [6] P. Antoniadis, I. Pikoulis, P. P. Filntisis, and P. Maragos. "[An audiovisual and contextual approach for categorical and continuous emotion recognition in-the-wild](#)," Proc. ICCV Workshop, 2021.
- [7] L. Wang et al. "Temporal segment networks: Towards good practices for deep action recognition." Proc ECCV 2016.
- [8] Kirillov, Alexander, et al. "Panoptic segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

V2. Three-dimensional Shape: Analysis, Modeling, Matching

References:

- [1] A. Bronstein, M. Bronstein, and R. Kimmel. [Numerical geometry of non-rigid shapes](#). Springer, 2008.
- [2] M. Breuss, A. Bruckstein and P. Maragos (Eds.), [Innovations for Shape Analysis: Models and Algorithms](#), Springer, 2013.
- [3] J. Sun, M. Ovsjanikov, L. Guibas, "[A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion](#)", Eurographics Symposium on Geometry Processing 2009.

V3. Graph-theoretic Methods for Clustering and Segmentation (extension to non-planar graphs)

References:

- [1] C. G. Bampis, P. Maragos and A. C. Bovik, "[Graph-Driven Diffusion and Random Walk Schemes for Image Segmentation](#)," IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 35-50, Jan. 2017.
- [2] C. Sakaridis, K. Drakopoulos and P. Maragos, "[Theoretical Analysis of Active Contours on Graphs](#)," SIAM J. Imaging Sciences, vol.10, no. 3, pp. 1475-1510, 2017.

V4. Sign-Language Understanding using Deep Learning

V4.1 Sign-Language Recognition using Deep Learning: Αναγνώριση χειρονομιών και συνεχών βίντεο νοηματικής γλώσσας με την χρήση βαθείων νευρωνικών δικτύων.

(Συνεργάτες: Αν. Καθ. Γ. Ποταμιάνος, Τμήμα ΗΜΜΥ, Παν/μιο Θεσσαλίας)

References:

- [1] S. Theodorakis, V. Pitsikalis and P. Maragos, “Dynamic–static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition”, *Image and Vision Computing*, 32 (8), pp.533–549, Aug. 2014.
- [2] A. Roussos, S. Theodorakis, V. Pitsikalis and P. Maragos, “[Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos](#)”, *J. Machine Learning Research*, 2013.
- [3] O. Koller, N.C. Camgoz, H. Ney, and R. Bowden, “Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos,” *IEEE Trans. PAMI*, 2020. DOI: 10.1109/TPAMI.2019.2911077
- [4] M. Parelli, K. Papadimitriou, G. Potamianos, G. Pavlakos and P. Maragos, “[Exploiting 3D Hand Pose Estimation in Deep Learning-based Sign Language Recognition from RGB Videos](#)”, *Proc. ECCV Workshops (ECCVW-2020) - Workshop on Sign Language Recognition, Translation and Production (SLRTP 2020)*, Aug. 2020.
- [5] M Parelli, K Papadimitriou, G Potamianos, G Pavlakos, P Maragos, “[Spatio-Temporal Graph Convolutional Networks for Continuous Sign Language Recognition](#)”, *Proc. ICASSP-2022*.
- [6] A. Kratimenos, G. Pavlakos and P. Maragos, “[Independent Sign Language Recognition With 3D Body, Hands, And Face Reconstruction](#)”, *Proc. ICASSP-2021*.
- [7] C. C. de Amorim, D. Macêdo and C. Zanchettin, “Spatial-Temporal Graph Convolutional Networks for Sign Language Recognition”, <https://arxiv.org/abs/1901.11164>, 2019.
- [8] K Papadimitriou, G Potamianos, G Sapountzaki, T Goulas, E Efthimiou, S -E Fotinea, P Maragos, “[Greek sign language recognition for an education platform](#)”, *Universal Access in the Information Society*, 2023.

(*) V4.2 Sign-Language Generation using Deep Learning: Automatic generation of sign-language gestures from text and neural photorealistic rendering of videos.

The primary goal of this project is to revolutionize Sign Language Synthesis technology by developing an automatic system for transforming spoken language text into hyper-realistic sign language gestures. Current technology employs 3D avatars to translate spoken language into sign language, but it lacks realism in appearance, movement, and neglects the critical non-manual signals such as facial expressions. This project aims to address these limitations, generating highly realistic videos featuring not just hands, but also accurate body, head, and facial motions during signing. Furthermore, the system could render any real person, from celebrities to historical figures, as the virtual sign language interpreter. The successful implementation of this system could usher in a new era of sign language synthesis that significantly improves user engagement and experience, with potential applications in education, museums, and safety information. This ambitious yet feasible project builds on recent advances in neural rendering for facial and head re-enactment, along with state-of-the-art methods for deep 3D reconstruction of hands and full body. While there have been a few recent attempts at photo-realistic sign language synthesis, they suffer from limitations in modeling accuracy and realism, which this project seeks to overcome. The primary focus of this project is the automatic generation of sign language gestures from text.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ, και Δρ. Αναστάσιος Ρούσσοσ, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ΙΤΕ)

References:

- [1] Saunders, B., Camgoz, N. C., & Bowden, R. (2022). Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5141-5151).
- [2] Forte, M. P., Kulits, P., Huang, C. H. P., Choutas, V., Tzionas, D., Kuchenbecker, K. J., & Black, M. J. (2023). Reconstructing Signing Avatars From Video Using Linguistic Priors. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR-2023)*.
- [3] Tze, C. O., Filntisis, P. P., Dimou, A. L., Roussos, A., & Maragos, P. (2023). “[Neural Sign Reenactor: Deep Photorealistic Sign Language Retargeting](#)”. *CVPRW Workshop AI4CC(2023)*

V5. 3D Computer Vision for Faces and Other Deformable Objects

(*) V5.1 3D face modeling/reconstruction and applications (photorealistic avatars, talking heads)

The human face is one of the most commonly-considered objects in Computer Vision and Graphics. Modelling and reconstructing the detailed 3D shape and dynamics of the human face has numerous applications, such as augmented reality, performance capture, computer games, visual effects, human-computer interaction, computer-aided craniofacial surgery, rehabilitation and research in psychology, to name a few. During the last years we have proposed state-of-the-art methods for 3D face modelling, 3D face reconstruction from in-the-wild images

and videos, facial expression recognition and photo-realistic synthesis of facial videos, see e.g. [1-5]. There can be various interesting projects related to the aforementioned state-of-the-art methods. These methods have yielded promising results, but there are many ways that they can be improved, extended, studied further or applied to different problems.

(Συνεργάτες: Δρ. Αναστάσιος Ρούσσοσ, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ΙΤΕ, και Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] M.C. Doukas, M.R. Koujan, V. Sharmanska, A. Roussos, and S. Zafeiriou, "[Head2Head++: Deep Facial Attributes Re-Targeting](#)", IEEE Transactions on Biometrics, Behavior, and Identity Science (IEEE T-BIOM), Volume 3, Issue 1, pp 31-43, January 2021.
- [2] M.R. Koujan, ..., A. Roussos, "[Real-time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity](#)", Proc. Int'l Conf. Automatic Face and Gesture Recognition (FG 2020).
- [3] M.R. Koujan, A. Roussos and S. Zafeiriou. "[DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation](#)", International Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [4] J. Booth, A. Roussos et al, "[3D reconstruction of in-the-Wild Faces in Images and Videos](#)", PAMI 2018.
- [5] J. Booth, A. Roussos, ..., S. Zafeiriou, "[Large Scale 3D Morphable Face Models](#)", IJCV 2017.
- [6] S. Ploumpis et al, "Towards a complete 3D morphable model of the human head", <https://arxiv.org/abs/1911.08008> (CVPR and PAMI reprint).

(*) V6. 6DoF Pose Estimation of Objects with Emphasis on Dense Implicit Embeddings.

Ανάπτυξη αλγορίθμων εντοπισμού 3D πόζας αντικειμένων με τεχνικές βαθιάς μάθησης. Συγκεκριμένα ο εντοπισμός πόζας θα γίνεται είτε από (monocular) RGB εικόνες είτε από RGB-D εικόνες και θα δοθεί έμφαση στην κωδικοποίηση της πληροφορίας πόζας με ερμηνεύσιμες αναπαραστάσεις σε επίπεδο pixel ([3]). Ιδιαίτερη προσοχή θα δοθεί στον εντοπισμό πόζας συμμετρικών αντικειμένων.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] Peng, Sida, et al. "Pvnet: Pixel-wise voting network for 6dof pose estimation." Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [2] He, Yisheng, et al. "Pvn3d: A deep point-wise 3d keypoints voting network for 6dof pose estimation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [3] Retsinas, George, Niki Efthymiou, and Petros Maragos. "[Mushroom Segmentation and 3D Pose Estimation From Point Clouds Using Fully Convolutional Geometric Features and Implicit Pose Encoding](#)." Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

(*) V7. Image Manipulation via Guiding Attention Maps for Vision-Language Models.

Μελέτη συστημάτων κοινής περιγραφής πληροφορίας κειμένου/όρασης, όπως CLIP[1], και δημιουργίας εικόνων από κείμενο, π.χ. Imagen[2], για την εξαγωγή συμπερασμάτων (explainable-AI) για την μορφή των χαρτών προσοχής (attention masks). Επέκταση τεχνικών επεξεργασίας εικόνων (image editing) μέσω εντολών σε μορφή κειμένου (text prompts) που βασίζονται σε attention λογικές (π.χ. [3]).

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.
- [2] Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." *Advances in Neural Information Processing Systems* 35 (2022): 36479-36494.
- [3] Mokady, Ron, et al. "Null-text inversion for editing real images using guided diffusion models." Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

(* V8. 3D Child Pose and Shape Estimation from images.

Most current methods for 3D Human Pose and Shape Estimation [1] rely on parametric body models that are representative of adult body shapes [2]. This reliance limits these methods, as they cannot accurately reconstruct the 3D shape and pose for children, leading to unsatisfactory results when applied to younger age groups. The goal of this project is to leverage recent advances in human body modeling [3,4] to develop techniques that can reconstruct 3D shapes across a wider variety of body shapes, including those of children and babies [5,6].

(Συνεργάτες: Επ. Καθ. Γιώργος Παυλάκος, UT Austin, Δρ. Νίκη Ευθυμίου)

References:

- [1] Goel, Shubham, et al. "[Humans in 4D: Reconstructing and tracking humans with transformers.](#)" Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
- [2] Loper, Matthew, et al. "[SMPL: A Skinned Multi-Person Linear Model.](#)" ACM Transactions on Graphics. 2015.
- [3] Hesse, Nikolas, et al. "[Learning and tracking the 3D body shape of freely moving infants from RGB-D sequences.](#)" IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019.
- [4] Sun, Yu, et al. "[Putting people in their place: Monocular regression of 3D people in depth.](#)" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [5] Patel, Priyanka, et al. "[AGORA: Avatars in geography optimized for regression analysis.](#)" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
- [6] Müller, Lea, et al. "[Generative proxemics: A prior for 3D social interaction from images.](#)" Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

(* V9. Morphological Filtering and Semantic Segmentation of Hyperspectral Images (Θ. Ροντογιάννης και Π. Μαραγκός)

Η Μαθηματική Μορφολογία αποτελεί ένα ισχυρό μεθοδολογικό πλαίσιο για την μη-γραμμική επεξεργασία και ανάλυση εικόνας, βασισμένο σε θεωρία και τελεστές συνόλων και lattices [1,2,3]. Τα μορφολογικά φίλτρα είναι μη-γραμμικοί τελεστές που έχουν αποδειχθεί ιδιαίτερα αποτελεσματικοί σε προβλήματα που σχετίζονται με γεωμετρικά χαρακτηριστικά μιας εικόνας. Αρχικά τα μορφολογικά φίλτρα αναπτύχθηκαν και χρησιμοποιήθηκαν σε δυαδικές και γκριζες εικόνες [1,2,3], και πιο πρόσφατα η θεωρία και εφαρμογή τους επεκτάθηκε σε έγχρωμες [4] και υπερφασματικές εικόνες [5,6]. Λόγω της πλούσιας χωρικής και φασματικής πληροφορίας και της πληθώρας των εφαρμογών τους, η επεξεργασία και ανάλυση υπερφασματικών εικόνων [7,8,9] έχει προσελκύσει το ενδιαφέρον της ερευνητικής κοινότητας τα τελευταία χρόνια. Στο πλαίσιο αυτό, ένα πολύ ενδιαφέρον πρόβλημα υπολογιστικής όρασης είναι αυτό της σημασιολογικής κατάτμησης (semantic segmentation) μιας υπερφασματικής εικόνας, κατά την οποία κάθε pixel της εικόνας ταξινομείται σε μια κατηγορία ή κλάση. Στη διπλωματική αυτή θα διερευνηθεί κατά πόσο κατάλληλα επιλεγμένα μορφολογικά φίλτρα μπορούν να αξιοποιηθούν αποδοτικά για την σημασιολογική κατάτμηση υπερφασματικών εικόνων. Αρχικά θα θεωρήσουμε ότι τα φίλτρα εφαρμόζονται απ' ευθείας στον φασματικό κύβο και στη συνέχεια θα διερευνήσουμε διάφορους μετασχηματισμούς μείωσης της διαστατικότητας των δεδομένων (π.χ., PCA, tensor decompositions), πριν την εφαρμογή των φίλτρων στις μετασχηματισμένες εικόνες. Τα αποτελέσματα που θα προκύψουν θα συγκριθούν με τα αποτελέσματα άλλων μεθόδων κατάτμησης από τη βιβλιογραφία, τόσο κλασικών τεχνικών, όσο και τεχνικών βαθιάς μάθησης.

- [1] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, 1982.
- [2] P. Maragos, "[Morphological Signal and Image Processing](#)", Book Chapter in *Digital Signal Processing Handbook*, V. Madisetti and D. Williams (Eds.), CRC Press, Boca Raton, Florida, 1998, , pp. 74:1-74:31.
- [3] P. Maragos, "[Morphological Filtering for Image Enhancement and Feature Detection](#)", Chapter 3.3 for *Image and Video Processing Handbook (2nd ed.)*, A. Bovik (ed.), Acad. Press, 2005
- [4] J. Angulo, "Morphological colour operators in totally ordered lattices based on distances: Application to image filtering, enhancement and analysis," *Computer Vision and Image Understanding*, vol. 107, issue 1-2, pp. 2007.
- [5] G. Noyel, J. Angulo, D. Jeulin, "[Morphological segmentation of hyperspectral images.](#)" arXiv preprint arXiv:2010.00853, 2020
- [6] S. Velasco-Forero, J. Angulo, "[Classification of hyperspectral images by tensor modeling and additive morphological decomposition.](#)" *Pattern Recognition*, vol. 46, issue 2, pp. 566-577, Feb. 2013.
- [7] S. Velasco-Forero and J. Angulo, "[Vector Ordering and Multispectral Morphological Image Processing](#)". In: *Advances in Low-Level Color Image Processing*, M. Celebi and B. Smolka (eds), Lecture Notes in Computational Vision and Biomechanics, vol 11, pp.223-239, Springer, 2014.

- [8] K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "[A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing](#)," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 585-599, Feb. 2012.
- [9] P.V. Giampouras, K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "[Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing](#)," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, issue 8, pp. 4775-4789, Aug. 2016.
- [10] S.D. Xenaki, K.D. Koutroumbas, A.A. Rontogiannis, "[Sparsity-aware possibilistic clustering algorithms](#)," *IEEE Transactions on Fuzzy Systems*, vol. 24, issue 6, pp. 1611-1626, Dec. 2016.

Επεξεργασία Ακουστικών σημάτων (π.χ. μουσικής, φωνής) ή Οπτικών σημάτων σε Πολυμεσικά περιβάλλοντα:

AVM1. Υπολογιστική επεξεργασία μουσικών σημάτων και γενικότερα μουσικής πληροφορίας
(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση και ΥΔ Χρήστος Γαρούφης, ΕΜΠ)

(* AVM1.1: Αυτόματη αναγνώριση και εντοπισμός μουσικών χαρακτηριστικών (auto-tagging).

Στόχος αυτής της διπλωματικής εργασίας είναι η εκπαίδευση νευρωνικών δικτύων που θα αναγνωρίζουν είτε το ηχητικό (πχ όργανο ή μουσικό είδος), είτε το σημασιολογικό (πχ συναίσθημα, καλλιτέχνης) περιεχόμενο ενός μουσικού κομματιού. Συγκεκριμένα, μπορεί να ακολουθηθεί είτε μία task-specific προσέγγιση, με εκπαίδευση σε ένα συγκεκριμένο σύνολο δεδομένων [1-3], είτε μία προσέγγιση αυτο-επιβλεπόμενης εκμάθησης αναπαραστάσεων, μέσω αντιθετικής μάθησης (self-supervised contrastive learning) [4-8] ή ισαλλοιώτης μάθησης (equivariant learning) [9-10]. Στην περίπτωση αυτή, μπορούν να χρησιμοποιηθούν δεδομένα τόσο από την ίδια (ακουστική) τροπικότητα, όσο και διαφορετικών τροπικοτήτων κατά τη διαδικασία εκπαίδευσης [11-12], όπως για παράδειγμα των αποσπασμάτων μουσικής και των αντίστοιχων στίχων. Παράλληλα, ενδιαφέρον παρουσιάζει η χρήση του διαχωρισμού μουσικών πηγών ως βοηθητικό πρόβλημα για την εκμάθηση κατάλληλων αναπαραστάσεων για προβλήματα ταξινόμησης [3,8].

Datasets:

- MagnaTagATune: <https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset>
- MTG-Jamendo: <https://github.com/MTG/mtg-jamendo-dataset>
- DALI: <https://github.com/gabolsgabs/DALI>

References:

- [1] Choi, K., G. Fazekas, M. Sandler, and K. Cho. "[Convolutional Recurrent Neural Networks for Music Classification](#)," in Proc. ICASSP 2017
- [2] Kratimenos, A., K. Avramidis, C. Garoufis, A. Zlatintsi, and P. Maragos. [Augmentation Methods On Monophonic Audio For Instrument Classification In Polyphonic Music](#), in Proc. EUSIPCO–2020.
- [3] Garoufis, C., A. Zlatintsi, and P. Maragos. "[Pre-Training Music Classification Models via Music Source Separation](#)," in Proc. EUSIPCO 2024
- [4] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.. [A simple framework for contrastive learning of visual representations](#), in Proc. ICML 2020
- [5] Saeed, A., Grangier, D., & Zeghidour, N. [Contrastive learning of general-purpose audio representations](#). In *Proc. ICASSP 2021*
- [6] Spijkervet, Janne, and John Ashley Burgoyne. "[Contrastive learning of musical representations](#)," in Proc. ISMIR 2021
- [7] Alonso-Jiménez, Pablo, Xavier Serra, and Dmitry Bogdanov. "[Music representation learning based on editorial metadata from discogs](#)," in Proc. ISMIR 2022
- [8] Garoufis, Christos, Athanasia Zlatintsi, and Petros Maragos. "[Multi-Source Contrastive Learning from Musical Audio](#)," in Proc. SMC 2023
- [9] Quinton, Elio, "[Equivariant Self-Supervision for Musical Tempo Estimation](#)", in Proc. ISMIR 2022
- [10] Riou, Alain, Stefan Lattner, Gaëtan Hadjeres, and Geoffroy Peeters. "[PESTO: Pitch Estimation with Self-Supervised Transposition-Equivariant Objective](#)," in Proc. ISMIR 2023
- [11] Manco, Ilaria, Emmanouil Benetos, Elio Quinton and Gyorgy Fazekas, "[Contrastive audio-language learning for music](#)," in Proc. ISMIR 2022
- [12] Avramidis, Kleanthis, Shanti Stewart, and Shrikanth Narayanan. "[On the Role of Visual Context in Enriching Music Representations](#)," in Proc ICASSP 2023

(*) **AVM1.2: Ανάπτυξη συστήματος εξατομικευμένων μουσικών προτάσεων**

Αντικείμενο αυτής της διπλωματικής εργασίας είναι η δημιουργία ενός συστήματος μουσικών προτάσεων (music recommendation) ή συνέχισης playlists (playlist continuation), βάσει είτε μουσικών χαρακτηριστικών ή μουσικής ομοιότητας [1-3], είτε μεθόδους collaborative filtering [4-5] σε στατιστικά ακρόασης, με αρχιτεκτονικές βασισμένες σε Variational Autoencoders [6-7] να πετυχαίνουν το state-of-the-art στην περίπτωση αυτή. Πιθανές επεκτάσεις των παραπάνω αφορούν την ενσωμάτωση/ανάπτυξη πολυεπίπεδων αρχιτεκτονικών VAEs ώστε να λαμβάνουν υπόψη την ιεραρχία των μεταδεδομένων (καλλιτέχνης, άλμπουμ, κομμάτι), καθώς και τη διερεύνηση της χρήσης ακουστικής πληροφορίας για την εξισορρόπηση των μουσικών προτάσεων προερχόμενων από μη-κυρίαρχες μουσικές κουλτούρες [8]. Επιπρόσθετα, μία δυνατότητα αφορά την προσθήκη εξατομικεύσεως των προτάσεων, μέσω μεθοδολογιών ενεργής μάθησης (active learning) [9], ή αποσύνδεσης (decoupling) των διαφόρων μουσικών παραμέτρων, όπως η χροιά, το είδος, ή το επαγόμενο συναίσθημα [3,10].

Datasets:

- Music4All: <https://sites.google.com/view/contact4music4all>
- MusAV: <https://mtg.github.io/musav-dataset>
- MPD: <https://research.atspotify.com/2020/09/the-million-playlist-dataset-remastered/>

References:

- [1] McFee, B., L. Barrington, and G. R. G. Lanckriet, "[Learning Content Similarity for Music Recommendation](#)", *IEEE Trans. on Audio, Speech & Language Processing* 20, no.8 (2012): 2207-2218
- [2] Huang, Q., A. Jansen, L. Zhang, D.P. Ellis, R.A. Saurous, and J. Anderson, "[Large-scale weakly-supervised content embeddings for music recommendation and tagging](#)". In *Proc ICASSP 2020*
- [3] Lee, J., Bryan, N. J., Salamon, J., Jin, Z., & Nam, J. . "[Disentangled multidimensional metric learning for music similarity](#)." In *Proc. ICASSP 2020*
- [4] W. Bendada, G. Salha-Galvan, T. Bouabça, and T. Cazenave, "[A Scalable Framework for Automatic Playlist Continuation on Music Streaming Services](#)". In *Proc. SIGIR 2023*
- [5] M. Volkovs, H. Rai, Z. Cheng, G. Wu, Y. Lu, and S. Sanner, "[Two-stage Model for Automatic Playlist Continuation at Scale](#)" In *Proc.of the ACM Recommender Systems Challenge 2018, 2018*.
- [6] V. Vančura and P. Kordík. "[Deep Variational Autoencoder with Shallow Parallel Path for Top-N Recommendation \(VASP\)](#)." In *Proc. ICANN 2021*
- [7] I. Shenbin, A. Alekseev, E. Tutubalina, V. Malykh, and S. I. Nikolenko. "[REC-VAE: A New Variational Autoencoder for Top-N Recommendations with Implicit Feedback](#)." In *Proc. WSDM 2020*
- [8] O. Lesota et al. "[Traces of Globalization in Online Music Consumption Patterns and Results of Recommendation Algorithms](#)". in *Proc. ISMIR 2022*.
- [9] Cañón, J. S. G., Cano, E., Yang, Y. H., Herrera, P., & Gómez, E. . "[Let's agree to disagree: Consensus Entropy Active Learning for Personalized Music Emotion Recognition](#)." in *Proc. ISMIR 2021*
- [10] Wang, X., Chen, H., Zhou, Y., Ma, J., & Zhu, W. (2022). "[Disentangled representation learning for recommendation](#)". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 408-424.

(*) **AVM1.3: Διαχωρισμός μουσικών πηγών**

Αντικείμενο αυτής της διπλωματικής εργασίας αποτελεί το πρόβλημα του διαχωρισμού μουσικών πηγών (music source separation), της απομόνωσης δηλαδή των φωνητικών, ή των επιμέρους οργάνων, που αποτελούν ένα μουσικό κομμάτι. Το τελευταίο χρονικό διάστημα παρατηρείται μία τάση αντικατάστασης, ή ενίσχυσης, των παραδοσιακών, state-of-the-art αρχιτεκτονικών [1-4] με καθαρά παραγωγικές (generative) προσεγγίσεις, όπως αυτή των diffusion models [5-6]. Μία άλλη ενεργή πιθανή κατεύθυνση του προβλήματος είναι αυτή του query-based διαχωρισμού πηγών, όπου στόχος είναι ο διαχωρισμός μίας πηγής-στόχου από ένα ηχητικό μίγμα [7-8], ενώ με την πρόσφατη άνθηση των προ-εκπαιδευμένων μοντέλων [9] για προβλήματα μουσικής ταξινόμησης, παρουσιάζει ενδιαφέρον η προσαρμογή τους στο πρόβλημα του διαχωρισμού μουσικών πηγών [10-11].

Datasets:

- musdb18: <https://sigsep.github.io/datasets/musdb.html>
- URMP: <https://labsites.rochester.edu/air/projects/URMP.html>
- MoisesDB: <https://github.com/moises-ai/moises-db>

References:

- [1] Kong, Q., Cao, Y., Liu, H., Choi, K., & Wang, Y. [Decoupling magnitude and phase estimation with deep resnet for music source separation](#), in Proc. ISMIR 2021
- [2] Défossez, A.. [Hybrid spectrogram and waveform source separation](#), in Proc. MDX Workshop 2021
- [3] Papantonakis, P., Garoufis, C., & Maragos, P. [Multi-band Masking for Waveform-based Singing Voice Separation](#). In Proc. EUSIPCO 2022
- [4] Luo, Yi, and Jianwei Yu. "[Music Source Separation With Band-Split RNN](#)." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023).
- [5] Plaja-Roglans, G., Miron, M., & Serra, X. [A diffusion-inspired training strategy for singing voice extraction in the waveform domain](#), in Proc. ISMIR 2022
- [6] Mariani, G., Tallini, I., Postolache, E., Mancusi, M., Cosmo, L., & Rodolà, E. [Multi-source diffusion models for simultaneous music generation and separation](#). *arXiv preprint arXiv:2302.02257*.
- [7] Lin, L., Kong, Q., Jiang, J., & Xia, G. [A unified model for zero-shot music source separation, transcription and synthesis](#), in Proc. ISMIR 2021
- [8] Kong, Q., Chen, K., Liu, H., Du, X., Berg-Kirkpatrick, T., Dubnov, S., & Plumbley, M. D.,. [Universal Source Separation with Weakly Labelled Data](#). *arXiv preprint arXiv:2305.07447*.
- [9] Li, Y., R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao et al. "[MERT: Acoustic Music Understanding Model with Large-Scale Self-Supervised Training](#)", in Proc. ICLR 2024
- [10] Chen, Z., Kanda, N., Wu, J., Wu, Y., Wang, X., Yoshioka, T., Li, J., Sivasankaran, S. and Eskimez, S.E., "[Speech Separation with Large-Scale Self-Supervised Learning](#)", in Proc. ICASSP 2023
- [11] Chen, K., Wichern, G., Germain, F. G., & Le Roux J., "[PaG-HuBERT: Self-Supervised Music Source Separation Via Primitive Auditory Clustering And Hidden-Unit BERT](#)" in Proc. ICASSP-W 2023

AVM2. Audio-Visual Multi-tasking Network: Action Recognition, Saliency Estimation and Video Summarization: Οπτικοακουστικό δίκτυο που πραγματοποιεί ταυτόχρονα αναγνώριση δράσεων, εκτίμηση του saliency και περίληψη

References:

- [1] P. Koutras and P. Maragos, "[SUSiNet: See, Understand and Summarize It](#)", *Proc. CVPR Workshop* 2019.
- [2] A. Tsiami, P. Koutras and P. Maragos, "[STAViS: Spatio-Temporal AudioVisual Saliency Network](#)", *Proc. CVPR 2020*.
- [3] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, [Multimodal Saliency and Fusion for Movie Summarization based on Aural, Visual, and Textual Attention](#), *IEEE Trans. Multimedia*, Nov. 2013.
- [4] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos and P. Maragos, "[COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization](#)", *EURASIP Journal on Image and Video Processing* (2017) 2017:54.

AVM3. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα AM-FM και αλγορίθμους πολυζωνικής ενεργειακής αποδιαμόρφωσης, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "[Energy Separation in Signal Modulations with Application to Speech Analysis](#)", *IEEE Trans. Signal Processing*, vol.41, no.10, pp.3024-3051, Oct. 1993.
- [2] A. Potamianos and P. Maragos, "[Speech Processing Applications Using an AM-FM Modulation Model](#)", *Speech Communication*, vol.28, no.3, pp.195-209, July 1999.

- [3] D. Dimitriadis, P. Maragos, and A. Potamianos, "[Robust AM-FM Features for Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.12, no.9, pp.621-624, Sep. 2005.
- [4] J. H.L. Hansen and S. Patil, [Speech Under Stress: Analysis, Modeling and Recognition](#), Springer LNAI 4343, 2007.
- [5] D. Dimitriadis, P. Maragos and A. Potamianos, "[On the Effects of Filterbank Design and Energy Computation on Robust Speech Recognition](#)", *IEEE Trans. Audio, Speech and Language Processing*, vol.19, pp.1504-1516, Aug.2011.
- [6] A. Zlatintsi and P. Maragos, "[Comparison of Different Representations Based on Nonlinear Features for Music Genre Classification](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [7] T. Chaspari, D. Dimitriadis and P. Maragos, "[Emotion Classification of Speech Using Modulation Features](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [8] I. Rodomagoulakis and P. Maragos, "[On the Improvement of Modulation Features Using Multi-Microphone Energy Tracking for Robust Distant Speech Recognition](#)", *Proc. EUSIPCO-2017*, Kos, Greece, Aug. 2017.
- [9] H. B. Sailor, M. R. Kamble and H. A. Patil, "[Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection](#)", in *Proc. Interspeech 2018*.

AVM4. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα από φράκταλς και χάος, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos and A. Potamianos, "[Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition](#)", *Journal of Acoustical Society of America*, vol.105 (3), pp.1925--1932, March 1999.
- [2] V. Pitsikalis and P. Maragos, "[Filtered Dynamics and Fractal Dimensions for Noisy Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.13, no.11, pp.711-714, Nov. 2006..
- [3] V. Pitsikalis and P. Maragos, "[Analysis and Classification of Speech Signals by Generalized Fractal Dimension Features](#)", *Speech Communication*, vol.51, no.12, pp.1206-1223, Dec. 2009.
- [4] A. Zlatintsi and P. Maragos, "[Multiscale Fractal Analysis of Musical Instrument Signals with Application to Recognition](#)", *IEEE Trans. Audio, Speech and Language Processing*, vol.21, no.4, pp.737-748, Apr. 2013.
- [5] Karnele López-de-Ipina et al., "[Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach](#)", *Computer Speech and Language* 30, pp.43–60, 2015.
- [6] E. Tzinis, G. Paraskevopoulos, C. Baziotis, A. Potamianos, "[Integrating Recurrence Dynamics for Speech Emotion Recognition](#)", in *Proc. Interspeech 2018*.
- [7] K. Avramidis, A. Zlatintsi, C. Garoufis, and P. Maragos, "[Multiscale Fractal Analysis on EEG Signals for Music-Induced Emotion Recognition](#)", *Proc. EUSIPCO, 2021*.
- [8] W. Gilpin, "Deep Reconstruction of Strange Attractors from Time Series", NIPS 2020, [arXiv:2002.05909 \(2020\)](#).

(* AVM5. Ανίχνευση βιοακουστικών γεγονότων με few-shot learning

Η ανίχνευση και ανάλυση βιοακουστικών σημάτων από φωνήματα θηλαστικών στη φύση είναι σημαντική για την μελέτη των πληθυσμών και της συμπεριφοράς τους ενώ ταυτόχρονα μπορεί να αποτελέσει ένα πολύτιμο εργαλείο για την προστασία απειλούμενων ειδών. Ο σκοπός της διπλωματικής εργασίας είναι η ανάπτυξη αποτελεσματικών αλγορίθμων, ικανών να κάνουν ταυτοποίηση και κατηγοριοποίηση φωνημάτων θηλαστικών στα χρονικά διαστήματα που αυτά συμβαίνουν με περιορισμένο αριθμό επισημειωμένων δειγμάτων. Συγκεκριμένα, θα χρησιμοποιηθεί το σύνολο δεδομένων του διαγωνισμού DCASE ([Detection and Classification of Acoustics Scenes and Events - task 5](#)). Στο σύνολο δεδομένων εκπαίδευσης (development set), τα δεδομένα είναι πλήρως επισημειωμένα με αναφορά στα χρονικά διαστήματα που ανιχνεύονται ήχοι (POS) ή δεν ανιχνεύονται (NEG) ή υπάρχει αβεβαιότητα (UNK) για κάθε κλάση ενώ το σύνολο αξιολόγησης (evaluation set) του αλγόριθμου περιλαμβάνει επισημειώσεις μόνο για τις πέντε πρώτες ανιχνεύσεις σε κάθε αρχείο. Το σύστημα βιοακουστικής ανίχνευσης που θα αναπτυχθεί στα πλαίσια της διπλωματικής, θα γενικεύει χαρακτηριστικά των φωνημάτων που θα εξάγει από τις 5 αναφορές ανά ηχητικό αρχείο για να ανιχνεύει επακόλουθα στιγμιότυπα των φωνημάτων στο υπόλοιπο αρχείο.

(Συνεργάτες: ΥΔ Δημήτρης Μακρόπουλος)

References:

- [1] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, [Matching Networks for One Shot Learning](#), Advances in Neural Information Processing Systems, 2016.

- [2] J. Snell, K. Swersky, and R. Zemel, [Prototypical Networks for Few Shot Learning](#), Advances in Neural Information Processing Systems, 2017.
- [3] L. You, E. P. Coyotl, S. Gunturu and M. Van Segbroeck, [Transformer-Based Bioacoustic Sound Event Detection on Few-Shot Learning Tasks](#), in Proc. IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.
- [4] J Liang, I Nolasco, B Ghani, H Phan, E Benetos, D Stowell, [Mind the Domain Gap: a Systematic Analysis on Bioacoustic Sound Event Detection](#), Eusipco, 2024.
- [5] E.Cakir, S.Adavanne, G.Parascandolo, K.Drossos, T.Virtanen, [Convolutional recurrent neural networks for bird audio detection](#), Eusipco, 2017.
- [6] D.Makropoulos, A.Tsiami, A.Prospathopoulos, D. Kassis, A. Frantzis, G. Piperakis, P. Maragos, [Convolutional Recurrent Neural Networks for the Classification of Cetacean Bioacoustic Patterns](#), in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

Συστήματα, Δίκτυα, Γράφοι, Optimization & Learning:

S1. Μη-γραμμικά δυναμικά συστήματα που χρησιμοποιούν max-plus άλγεβρα και finite-state automata με εφαρμογές σε ένα πρόβλημα από τις περιοχές ανίχνευσης, βελτιστοποίησης, δικτύων, ελέγχου, θεωρία γράφων, κ.ά.

References:

- [1] R. Cuninghame-Green, [Minimax Algebra](#), Springer-Verlag, New York, 1979.
- [2] P. Butkovič. [Max-linear Systems: Theory and Algorithms](#). Springer, 2010.
- [3] P. Maragos, [“Dynamical Systems on Weighted Lattices: General Theory”](#), *Mathematics of Control, Signals and Systems*, 29:21, 2017.
- [4] I. Kordonis, P. Maragos, G. P. Papavassilopoulos, [“Stochastic Stability in Max-Product and Max-Plus Systems with Markovian Jumps”](#), *Automatica* 92 (2018) 123–132.
- [5] S. Gaubert, [“Tropical convexity and games: regression, separation, and beyond”](#), CaLISTA Workshop: Geometry Informed Machine Learning, Paris, 2024.

(* **S2. Sparsity and Max-plus Algebra, Statistical Analysis of Tropical Estimators**

(Συνεργάτες: Νίκος Τσιλιβής, Courant Institute of New York)

- [1] A. Tsiamis and P. Maragos, [“Sparsity in Max-plus Algebra”](#), *Discrete Events Dynamic Systems*, 29:163–189, May 2019. <https://doi.org/10.1007/s10626-019-00281-1>
- [2] N. Tsilivis, A. Tsiamis and P. Maragos, [“Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression”](#), *Proc. ICASSP-2021*, June 2021.
- [3] N. Tsilivis, A. Tsiamis and P. Maragos, [“Sparse Approximate Solutions to Max-Plus Equations”](#), In: *Discrete Geometry and Mathematical Morphology: DGMM-2021*, J. Lindblad et al. (Eds), LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39
- [4] N Tsilivis, A Tsiamis, P Maragos, [“Toward a Sparsity Theory on Weighted Lattices”](#), *J. Mathematical Imaging and Vision*, 2022.

S3. Signal Processing on Graphs

References:

- [1] A. Sandryhaila and J. Moura, [“Discrete Signal Processing on Graphs”](#), *IEEE Trans. Signal Processing*, Apr. 2013.
- [2] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, [“The Emerging Field of Signal Processing on Graphs”](#), *IEEE Signal Processing Magazine*, May 2013.
- [3] G. Giannakis, Y. Shen and G. Karanikolas, [“Topology Identification and Learning Over Graphs: Accounting for Nonlinearities and Dynamics”](#), *Proc. IEEE*, vol.106, May 2018.

Machine Learning:

L1. Machine Learning, Deep Neural Nets, and Geometry

(* L1.1 Geometry, Training, Applications of DNNs with PWL Activations for Classification or Regression Problems

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

- Τα τροπικά μαθηματικά μπορούν να δώσουν κλειστές λύσεις για piecewise-linear approximations [4]. Μετέπειτα εργασίες, επεκτείνουν το framework ώστε να γίνεται sparse approximation των πολυδιάστατων επιφανειών [9], ή προσέγγισή τους με εναλλασσόμενη βελτιστοποίηση [13]. Στην περίπτωση του sparse approximation, η επιλογή των στοιχείων γίνεται από μια δοσμένη “βάση” πιθανών στοιχείων. Στο προκειμένο θέμα εξετάζουμε τρόπους να μάθουμε από τα δεδομένα τη βάση των πιθανών στοιχείων για το sparse approximation.
- Ενώ τα τροπικά μαθηματικά μπορούν να χρησιμοποιηθούν για piecewise-linear approximations [4], τα approximations είναι πάντα convex. Κάτι τέτοιο περιορίζει την ευρεία εφαρμογή των ανεπτυγμένων μεθόδων, και παράλληλα δεν επιτρέπει το approximation με αυθαίρετη ευκρίνεια. Σύγχρονες εργασίες προσεγγίζουν τα approximation με tropical rational functions [13], που βοηθά να αρθούν αυτοί οι περιορισμοί. Σε αυτό το θέμα εξετάζουμε περαιτέρω τρόπους για να αρθεί αυτός ο περιορισμός (π.χ. difference of convex functions) ή τρόπους να χαρακτηριστεί το approximation error έναντι κάποιων βέλτιστων μεθόδων (π.χ. με upper και lower envelopes).

References:

- [1] V. Charisopoulos and P. Maragos, “[Morphological Perceptrons: Geometry and Training Algorithms](#),” *Proc. ISMM 2017*. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.
- [2] V. Charisopoulos and P. Maragos, “[A tropical approach to neural networks with piecewise linear activations](#)”, arXiv:1805.08749, 2018.
- [3] E. Theodosis and P. Maragos, “[Analysis of the Viterbi Algorithm Using Tropical Algebra and Geometry](#)”, Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC-18), Greece, June 2018.
- [4] P. Maragos and E. Theodosis, “[Multivariate Tropical Regression and Piecewise-Linear Surface Fitting](#)”, *Proc. 45th IEEE Int’l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020)*, May 2020.
- [5] G. Smyrnis and P. Maragos, “[Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation](#)”, *Proc. 37th Int’l Conf. on Machine Learning (ICML-2020)*, July 2020.
- [6] I. Kordonis and P. Maragos, “Revisiting Tropical Polynomial Division: Theory, Algorithms and Application to Neural Networks”, arXiv:2306.15157: <https://arxiv.org/abs/2306.15157>
- [7] P. Maragos and E. Theodosis, “[Tropical Geometry and Piecewise-Linear Approximation of Curves and Surfaces on Weighted Lattices](#)”, arXiv:1912.03891: <https://arxiv.org/pdf/1912.03891>
- [8] N. Dimitriadis and P. Maragos, “[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)”, *Proc. ICASSP-2021*, June 2021.
- [9] N. Tsilivis, A. Tsiamis and P. Maragos, “[Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression](#)”, *Proc. ICASSP-2021*, June 2021.
- [10] N. Tsilivis, A. Tsiamis and P. Maragos, “Sparse Approximate Solutions to Max-Plus Equations”, In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology*: DGMM-2021, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39
- [11] P. Maragos, V. Charisopoulos, E. Theodosis, “Tropical Geometry and Machine Learning”, *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>
- [12] K. Fotopoulos, P. Maragos and P. Misiakos, “TropNNC: Structured Neural Network Compression Using Tropical Geometry”, arXiv:2409.03945: <https://www.arxiv.org/abs/2409.03945>
- [13] A. Dunbar and L. Ruthotto, “Alternating minimization for regression with tropical rational functions”, *Algebraic Statistics*, 15 (1), pp. 85-111, 2024. arXiv:2305.20072: <https://arxiv.org/pdf/2305.20072>

(* L1.2 Theoretical Analysis of Deep Neural Networks using Tropical Geometry Tools

- Μελέτη τροπικών ιδιοτήτων νευρωνικών δικτύων, με έμφαση σε βαθιές αρχιτεκτονικές, CNNs, ResNets, ή Kolmogorov Arnold Networks (KANs).
- Χρήση τροπικών ιδιοτήτων για μελέτη συμπεριφοράς δικτύων σε προβλήματα μηχανικής μάθησης.
- Πρακτική εφαρμογή στα ανωτέρω προβλήματα (ενδεικτικό πρόβλημα: συμπίεση νευρωνικών δικτύων).

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

References:

- [1] V. Charisopoulos and P. Maragos, “[A tropical approach to neural networks with piecewise linear activations](https://arxiv.org/pdf/1805.08749)”, arXiv:1805.08749, 2018, <https://arxiv.org/pdf/1805.08749>.
- [2] L. Zhang, G. Naitzat, and L.-H. Lim. “[Tropical geometry of deep neural networks](https://arxiv.org/pdf/1805.07091)”, *Proc. 35th Int’l Conf. on Machine Learning (ICML-2018)*, July 2018, <https://arxiv.org/pdf/1805.07091>.
- [3] P. Maragos, V. Charisopoulos, E. Theodosis, “Tropical Geometry and Machine Learning”, *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>
- [4] G. Smyrnis and P. Maragos, “[Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation](https://arxiv.org/pdf/2007.07091)”, *Proc. 37th Int’l Conf. on Machine Learning (ICML-2020)*, July 2020.
- [5] I. Kordonis and P. Maragos, “Revisiting Tropical Polynomial Division: Theory, Algorithms and Application to Neural Networks”, *arXiv:2306.15157*: <https://arxiv.org/abs/2306.15157>
- [6] N. Dimitriadis and P. Maragos, “[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](https://arxiv.org/pdf/2106.07091)”, *Proc. IEEE ICASSP 2021*, June 2021.
- [7] P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, “[Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes](https://openreview.net/pdf?id=oiZJwC_fyS)”, *Proc. ICLR 2022*, April 2022 https://openreview.net/pdf?id=oiZJwC_fyS.
- [8] K. Fotopoulos, P. Maragos and P. Misiakos, “TropNNC: Structured Neural Network Compression Using Tropical Geometry”, *arXiv:2409.03945*: <https://www.arxiv.org/abs/2409.03945>
- [9] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T.Y. Hou, and M. Tegmark, 2024 “KAN: Kolmogorov-Arnold Networks,” *arXiv:2404.19756* <https://arxiv.org/pdf/2404.19756>.
- [10] M.-C. Brandenburg, G. Loho and G. Montufar, "The Real Tropical Geometry of Neural Networks": <https://arxiv.org/abs/2403.11871>.
- [11] G. Montufar et al., “[Sharp Bounds for the Number of Regions of Maxout Networks and Vertices of Minkowski Sums](https://arxiv.org/pdf/2206.07091)”, *SIAM J. Applied Algebra and Geometry*, vol. 6(4), pp.618-649, 2022.

(* L2. Deep Neural Network Compression.

Ελαχιστοποίηση χωρικών/χρονικών απαιτήσεων σύγχρονων νευρωνικών δικτύων: Μελέτη του προβλήματος συμπίεσης νευρωνικών μέσω magnitude pruning των βαρών ενός NN. Έμφαση σε συναρτήσεις κατωφλιοποίησης και στην ευστάθεια της pruning μάσκας για μεγάλα ποσοστά συμπίεσης και επέκταση σε group sparsity λογικές. Ενδεικτικές εναλλακτικές κατευθύνσεις: quantization, weight sharing.

Αξιολόγηση σε δημοφιλείς βάσεις ταξινόμησης εικόνων, όπως:

<https://paperswithcode.com/sota/image-classification-on-cifar-100>.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] A.G. Georgoulakis, G. Retsinas and P. Maragos, “[Feather: An Elegant Solution to Effective DNN Sparsification](https://arxiv.org/pdf/2306.07091)”, *British Machine Vision Conference (BMVC)*, 2023.
- [2] K. S. Tai, T. Tian & S. N. Lim, “Spartan: Differentiable sparsity via regularized transportation”, *NeurIPS*, 2022.
- [3] G. Retsinas, A. Elafrou, G. Goumas and P. Maragos. “[Weight Pruning via Adaptive Sparsity Loss](https://arxiv.org/pdf/2006.02768)”, arXiv preprint arXiv:2006.02768, 2020.
- [4] Y. Cheng, D. Wang, P. Zhou and T. Zhang, “[A survey of model compression and acceleration for deep neural networks](https://arxiv.org/pdf/1710.09282)”, arXiv preprint arXiv:1710.09282, 2017.
- [5] M. Rastegari, V. Ordonez, J. Redmon and A. Farhadi, “[Xnor-net: Imagenet classification using binary convolutional neural networks](https://arxiv.org/pdf/1607.01725)”, *ECCV*, 2016.

(* L3. Προβλήματα Προσέγγισης στην Τροπική Άλγεβρα

Η τροπική άλγεβρα και γεωμετρία μπορούν να χρησιμοποιηθεί για την αναπαράσταση σε συμπαγή μορφή και την ανάλυση πολλών νευρωνικών δικτύων (με κατά τμήματα γραμμική ενεργοποίηση π.χ. ReLU, Maxout, κλπ), καθώς και την περιγραφή συνήθων αλγορίθμων όπως ο Viterbi [1]. Επιπλέον, στην τροπική άλγεβρα, ορισμένα προβλήματα μπορούν να λυθούν πολύ αποδοτικά (π.χ. η λύση γραμμικού συστήματος είναι πολύ πιο σύντομη στην τροπική άλγεβρα από ότι στη συνήθη [2]). Η διπλωματική αυτή ασχολείται με προβλήματα προσέγγισης στην Τροπική Άλγεβρα, δηλαδή περιπτώσεις στις οποίες θέλουμε να προσεγγίσουμε μια τροπική απεικόνιση από δεδομένα. Σε αυτό περιλαμβάνονται ενδεικτικά ένα ή περισσότερα από τα παρακάτω:

- **Προβλήματα κυρτής/τροπικής παλινδρόμησης με outliers.** Υπάρχει αρκετή προηγούμενη δουλειά σε κυρτή/τροπική παλινδρόμηση ([3],[4],[5]). Στα γραμμικά μοντέλα υπάρχουν πολύ επιτυχημένοι αλγόριθμοι για γραμμική παλινδρόμηση με outliers (π.χ. iteratively reweighted least squares, RANSAC). Στόχος εδώ είναι να επεκταθούν/τροποποιηθούν οι αλγόριθμοι της γραμμικής παλινδρόμησης στην περίπτωση της κυρτής/τροπικής.

- **Παραγοντοποίηση/Προσέγγιση Πινάκων στην Τροπική Άλγεβρα.** Η προσέγγιση χαμηλής τάξης των πινάκων καθώς και η παραγοντοποίηση με πίνακες με μη αρνητικούς όρους (Nonnegative Matrix Factorization) έχουν βρει πολλές εφαρμογές, όπως για παράδειγμα η ανάλυση κειμένων, η συμπίεση και τα recommender systems. Έχει ενδιαφέρον να επεκταθεί η προσεγγιστική παραγοντοποίηση στην τροπική περίπτωση. Κάποιες προκαταρκτικές εργασίες είναι οι [6],[7],[8]. Στόχος της διπλωματικής θα είναι να προταθούν πιο αποτελεσματικοί αλγόριθμοι, και να επεκταθεί το πεδίο εφαρμογής σε περιπτώσεις που οι πίνακες που αναζητούμε έχουν επιπλέον ιδιότητες (π.χ. χαμηλή τάξη).

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

References:

- [1] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", Proceedings of the IEEE, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>
- [2] P. Butkovič. Max-linear systems: theory and algorithms. Springer Science & Business Media, 2010.
- [3] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," Optimization and Engineering, vol. 10, no. 1, pp. 1–17, 2009, https://www-leland.stanford.edu/~boyd/papers/pdf/cvx_pwl_fit.pdf
- [4] P. Maragos and E. Theodosis, "Tropical geometry and piecewise-linear approximation of curves and surfaces on weighted lattices," ArXiv preprint arXiv:1912.03891, 2019, <https://arxiv.org/pdf/1912.03891>
- [5] J. Hook, "Linear regression over the max-plus semiring: algorithms and applications," arXiv preprint arXiv:1712.03499, 2017, <https://arxiv.org/pdf/1712.03499>.
- [6] S. Karaev and P. Miettinen, "Capricorn: An algorithm for subtropical matrix factorization," in Proceedings of the 2016 SIAM international conference on data mining. SIAM, 2016, pp. 702–710, <https://cs.uef.fi/~pauli/papers/karaev16capricorn.pdf>
- [7] S. Karaev and P. Miettinen "Cancer: Another algorithm for subtropical matrix factorization," in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II 16. Springer, 2016, pp. 576–592
- [8] A. Omanovic, P. Oblak, and T. Curk, "Faststmf: Efficient tropical matrix factorization algorithm for sparse data," arXiv preprint arXiv:2205.06619, 2022. <https://arxiv.org/pdf/2205.06619>
- [9] I Kordonis, E Theodosis, G Retsinas, P Maragos, "[Matrix Factorization in Tropical and Mixed Tropical-Linear Algebras Inproceedings](#)", Proc. ICASSP 2024, Seoul, Korea, 2024,

(*) L4. Μοντέλα Διάχυσης για Χρονοσειρές και Εφαρμογές

Τα τελευταία χρόνια υπάρχει μεγάλο ενδιαφέρον για generative μοντέλα. Μία από τις πιο επιτυχημένες κατηγορίες είναι τα μοντέλα διάχυσης [1-3]. Τα μοντέλα διάχυσης έχουν πρόσφατα επεκταθεί σε δεδομένα χρονοσειρών [4,5]. Αυτό επιτρέπει να δειγματοληπτούμε πιθανές μελλοντικές τροχιές από χρονοσειρές. Μερικά παραδείγματα χρονοσειρών είναι η ζήτηση ενός προϊόντος (π.χ. ενέργειας), η χρήση δεδομένων, η κίνηση στους δρόμους, οι μετρήσεις διαφόρων αισθητήρων κ.λπ. Στη διπλωματική αυτή προτείνουμε τη μελέτη generative μοντέλων για χρονοσειρές με έμφαση στη χρήση τους για βελτιστοποίηση. Ένα συγκεκριμένο παράδειγμα είναι η αποθήκευση ηλεκτρικής ενέργειας. Προτείνουμε την κατασκευή ενός μοντέλου που να δημιουργεί δείγματα από πιθανές μελλοντικές τροχιές τιμών και προσαρμόζουμε τις τεχνικές για την κατασκευή scenario trees (για τη χρήση τους στη βελτιστοποίηση π.χ. [6]). Αυτό μας επιτρέπει να βελτιστοποιήσουμε τη διαχείριση της αποθήκευσης της ενέργειας. Σαν σύγκριση μπορούμε να έχουμε το όφελος που θα προέκυπτε αν είχαμε γνώση της μελλοντικής τιμής.

(Συνεργάτες: Επ. Καθ. Ιωάννης Κορδώνης, ΕΜΠ)

References:

- [1] Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics." Proc. International Conference on Machine Learning, 2015 <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- [2] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851 <https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf>.
- [3] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." arXiv preprint arXiv:2011.13456 (2020) <https://openreview.net/pdf/ef0eadbe07115b0853e964f17aa09d811cd490f1.pdf>
- [4] Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Proc. International Conference on Machine Learning (ICML 2021), <https://proceedings.mlr.press/v139/rasul21a/rasul21a.pdf>.
- [5] Lin, Lequan, et al. "Diffusion models for time series applications: A survey." arXiv preprint arXiv:2305.00624 (2023).
- [6] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics, 2021.

[7] Zhang, Xinjing, et al. "Arbitrage analysis for different energy storage technologies and strategies." *Energy Reports* 7 (2021): 8198-8206.

[8] Kordonis, Ioannis, Alexandros C. Charalampidis, and Pierre Haessig. "Optimal operation of a grid-connected battery energy storage system over its lifetime." *Optimal Control Applications and Methods* 44.2 (2023): 739-757.

(* L5. Morphological (Max-plus) Networks for Classification or Regression Problems

Since 2012, there has been an explosion of machine learning research in recent years, which has now provided models in the form of deep neural networks (NNs) with near-human or even human performance. Recent advances indicate a shift towards nonlinearity, but the "multiply-accumulate" (= linear) operations of the perceptron are still ubiquitous.

Our Questions:

- Are dot products and convolutions the only biologically plausible models of neuronal computation?
- Can we use results and tools from nonlinear (in particular max-plus algebra and tropical geometry) mathematics to reason about complexity and dimension of deep learning models in current literature?
- Can we replace the "linear neurons" with nonlinear operations to achieve better performance?

Several research groups world-wide, inspired by the capabilities of nonlinear image analysis systems using morphological systems, have developed Morphological Neural Nets (MNNs), i.e. networks whose nodes perform max-plus operations. These NNs with morphological nodes can train faster (with a smoothed version of gradient descent and backpropagation) and have similar or better performance than NNs with linear nodes. Recently there is a growing interest for research on these nonlinear networks.

Στην διπλωματική αυτή προτείνουμε την μελέτη σύγχρονων MNNs με έναν από τους ακόλουθους στόχους: 1) καλύτερους αλγορίθμους για την εκπαίδευση τους, 2) βαθιές αρχιτεκτονικές, 3) εφαρμογές σε μεγαλύτερα datasets.

References:

[1] V. Charisopoulos and P. Maragos, "[Morphological Perceptrons: Geometry and Training Algorithms](#)," *Proc. ISMM* 2017. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.

[2] N. Dimitriadis and P. Maragos, "[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)," *Proc. IEEE ICASSP* 2021.

[3] P.-F. Yang and P. Maragos, "[Min-Max Classifiers: Theory, Design, and Application](#)", *Pattern Recognition*, 28 (6), pp.879-899, 1995.

[4] L. F.C. Pessoa and P. Maragos, "[Hybrid Neural Networks with Morphological/Rank/Linear Nodes: A Unifying Framework with Application to Handwritten Character Recognition](#)", *Pattern Recognition*, vol.33, pp.945-960, 2000.

[5] R. de A. Araújo, A. L.I. Oliveira and S. Meira, "[A morphological neural network for binary classification problems](#)", *Engineering Applications of Artificial Intelligence*, vol.65, pp.12-28, 2017.

[6] G. Franchi and A. Fehri and A. Yao, "[Deep Morphological Networks](#)", *Pattern Recognition*, 102, 107246, 2020.

[7] S. Velasco-Forero and J. Angulo, "[MorphoActivation: Generalizing ReLU Activation Function by Mathematical Morphology](#)", *Proc. DGMM 2022, LNCS*, vol. 13493, pp.449-461, Springer, 2022. Also <https://arxiv.org/abs/2207.06413>.

[8] M. E. Valle, "[Bridging Vector-Valued Mathematical Morphology and Deep Learning](#)", *CaLISTA Workshop: Geometry Informed Machine Learning*, Paris, 2024.

[9] J. G. Angulo, "[Lattice theory and algebraic models for deep learning based on mathematical morphology](#)", *CaLISTA Workshop: Geometry Informed Machine Learning*, Paris, 2024.

(* L6. Model-data-driven hyperspectral unmixing. (©. Ποντογιάννης)

Η πολυδιάστατη απεικόνιση με τη χρήση υπερφασματικών αισθητήρων (hyperspectral sensors) και η επεξεργασία και ανάλυση των υπερφασματικών δεδομένων που προκύπτουν συναντάται σε ένα ευρύτατο φάσμα εφαρμογών τα τελευταία χρόνια. Ένα ενδιαφέρον πρόβλημα στο πεδίο αυτό είναι αυτό του φασματικού διαχωρισμού υπερφασματικών εικόνων (hyperspectral unmixing), που αποτελεί ουσιαστικά ένα είδος προβλήματος τυφλού διαχωρισμού πηγών (blind source separation). Ο φασματικός διαχωρισμός αποσκοπεί τόσο στην αναγνώριση, με βάση τις φασματικές τους υπογραφές, των αντικειμένων ή υλικών που εμφανίζονται στην περιοχή απεικόνισης, όσο και στην περιγραφή της κατανομής των υλικών πάνω στην εικόνα.

Παραδοσιακά το πρόβλημα αυτό προσεγγίζεται με κατάλληλη μαθηματική μοντελοποίηση του μηχανισμού παραγωγής των υπερφασματικών δεδομένων και στην ανάπτυξη τεχνικών εκτίμησης των παραμέτρων του μοντέλου γέννησης των δεδομένων (model-based approach). Αξίζει να σημειωθεί ότι σημαντικές βελτιώσεις στις κλασσικές τεχνικές έχουν επιτευχθεί με την αξιοποίηση φυσικών περιορισμών που ικανοποιούν τα

δεδομένα, όπως μη αρνητικότητα (nonnegativity), αραιότητα (sparsity), χαμηλή τάξη (low-rank), κλπ. Πρόσφατα, η χρησιμοποίηση τεχνικών βαθιάς μάθησης έχει δώσει νέα ώθηση στο πεδίο του φασματικού διαχωρισμού υπερφασματικών εικόνων. Στο πλαίσιο της εργασίας αυτής, βασικός στόχος είναι η ανάπτυξη νέων αποδοτικών τεχνικών φασματικού διαχωρισμού, συνδυάζοντας την εμπειρία που έχει αποκτηθεί σε κλασικές μεθόδους με τις νέες ιδέες εξαγωγής γνώσης από τα δεδομένα που βασίζονται σε αρχιτεκτονικές και αλγορίθμους βαθιάς μάθησης.

References:

- [1] K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "[A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing](#)," IEEE Transactions on Signal Processing, vol. 60, no. 2, pp. 585-599, Feb. 2012.
- [2] P.V. Giampouras, K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "[Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing](#)," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, issue 8, pp. 4775-4789, Aug. 2016.
- [3] C. Tsinos, A.A. Rontogiannis, K. Berberidis, "[Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained nonnegative matrix factorization](#)," IEEE Transactions on Computational Imaging, vol. 3, issue 2, pp. 160-174, June 2017.
- [4] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "[Hyperspectral unmixing via deep convolutional neural networks](#)," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 11, pp. 1755-1759, 2018.
- [5] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "[Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing](#)," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 9, pp. 1427-1431, 2018.
- [6] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "[Deep generative endmember modeling: An application to unsupervised spectral unmixing](#)," IEEE Transactions on Computational Imaging, vol. 6, pp. 374-384, 2020.
- [7] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "[Misticnet: Minimum simplex convolutional network for deep hyperspectral unmixing](#)," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022.

(*) L7. Rank-revealing NMF for mutational signature separation in human cancer genomes (© Ροντογιάννης)

Η μέθοδος NMF (nonnegative matrix factorization) βασίζεται στη διάσπαση ενός μητρώου δεδομένων με μη αρνητικά στοιχεία σε δύο μητρώα-παράγοντες που έχουν επίσης μη αρνητικά στοιχεία [1]. Η μέθοδος αυτή έχει χρησιμοποιηθεί αποτελεσματικά σε μια ευρύτατη γκάμα διαφορετικών εφαρμογών [2], ιδιαίτερα σε περιπτώσεις που το φυσικό πρόβλημα μπορεί να μοντελοποιηθεί μαθηματικά ως ένα πρόβλημα τυφλού διαχωρισμού πηγών (blind source separation). Μια πολύ ενδιαφέρουσα εφαρμογή της μεθόδου NMF παρουσιάστηκε πρόσφατα σε μια μελέτη [3], που διήρκησε περισσότερο από 10 χρόνια με συμμετοχή ερευνητών από όλο τον κόσμο και αποσκοπούσε στην αποκωδικοποίηση και κατανόηση της σύνθεσης των καρκινικών γονιδιωμάτων του ανθρώπου. Στη μελέτη αυτή η μέθοδος NMF αξιοποιήθηκε για τον διαχωρισμό και την ταυτοποίηση των μεταλλακτικών υπογραφών (mutational signatures) που συνυπάρχουν και πιθανόν επικαλύπτονται στα καρκινικά γονιδιώματα, και έδωσε τη δυνατότητα αντιστοίχισης διαφόρων συνδυασμών υπογραφών με διάφορους τύπους καρκίνου. Όπως αναφέρεται στο [3], ένα στοιχείο που δυσχεραίνει τη διαδικασία αυτή είναι το γεγονός ότι ο αριθμός των μεταλλακτικών υπογραφών σε κάθε γονιδίωμα είναι άγνωστος. Στη διπλωματική αυτή θα διερευνήσουμε την αποτελεσματικότητα και αποδοτικότητα τεχνικών NMF που προτάθηκαν πρόσφατα [4],[5], οι οποίες πέρα από την παραγοντοποίηση του αρχικού μητρώου δεδομένων έχουν τη δυνατότητα να εκτιμούν και την τάξη του μοντέλου, που στην προκειμένη περίπτωση είναι ο αριθμός των μεταλλακτικών υπογραφών. Οι τεχνικές αυτές θα εφαρμοστούν σε πραγματικά δεδομένα γονιδιωμάτων από το [3] και θα συγκριθούν με άλλες NMF τεχνικές από την πρόσφατη βιβλιογραφία [6].

References:

- [1] D. Lee, H. Seung, "[Learning the parts of objects by non-negative matrix factorization](#)," Nature, vol. 401, no. 6755, pp. 788-791, Oct. 1999.
- [2] X. Fu, et. al., "[Nonnegative matrix factorization for signal and data analytics](#)," IEEE SP Magazine, pp. 59-80, March 2019.
- [3] L.B. Alexandrov, et. al., "[The repertoire of mutational signatures in human cancer](#)," Nature, <https://doi.org/10.1038/s41586-020-1943-3>, Feb. 2020.
- [4] P.V. Giampouras, A.A. Rontogiannis, K.D. Koutroumbas, "[A projected Newton type algorithm for nonnegative matrix factorization with model order selection](#)," in Proceedings of IEEE ICASSP, 2019.
- [5] L. Sfountouris, "[Hyperspectral image denoising and unmixing using variational Bayes and deep image prior](#)," Diploma Thesis, NTUA, Mar. 2024.
- [6] R. Nasser, Y.C. Eldar, R. Sharan, "[Deep unfolding for non-negative matrix factorization with application to mutational signature analysis](#)," Journal of Computational Biology, vol. 29, no. 1, Jan. 2022.

Επεξεργασία κειμένων και φυσικής γλώσσας:

T1. Learning from Structure: Annotation-free Text Recognition.

Χρήση της κατανομής λέξεων κειμένου για εκπαίδευση συστήματος αναγνώρισης χειρόγραφων λέξεων κειμένου χωρίς δεδομένα επισημείωσης. Συγκεκριμένα έχουμε 2 χώρους: τον χώρο οπτικών περιγραφητών από μεμονωμένες εικόνες χειρόγραφων λέξεων και τον χώρο των λέξεων σε μορφή γραμματοσειρών, όπου λέξεις με μικρές διαφορές, π.χ. έναν χαρακτήρα διαφορά: “true” και “tree”, πρέπει να βρίσκονται κοντά. Επιθυμούμε να ταιριάξουμε αυτούς τους δύο χώρους βασιζόμενοι στην χρήση της συχνότητας εμφάνισης συγκεκριμένων λέξεων (λέξεις όπως “and”, “the” είναι συχνές σε οποιοδήποτε κείμενο) αλλά και της εσωτερικής δομής τους (π.χ. κοινοί χαρακτήρες) για να περιγράψει μονοσήμαντα η κατανομή των λέξεων. Στην συνέχεια θα γίνει προσπάθεια αντιστοίχισης γραμματοσειρών λέξεων με περιγραφητές εικόνων λέξεων. Συνολικά το συγκεκριμένο πρόβλημα είναι μια εφαρμογή domain adaptation (π.χ. optimal transport).

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] G. Retsinas, G. Sfikas and P. Maragos, [“From Seq2Seq recognition to handwritten word embeddings”](#), BMVC, 2020.
- [2] Xu, Hongteng, et al. "Gromov-wasserstein learning for graph matching and node embedding." International conference on machine learning. PMLR, 2019.
- [3] Ding, Ning, et al. "Source-free domain adaptation via distribution estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

T2. Επεξεργασία κειμένων και φυσικής γλώσσας με αλγεβρικές και γεωμετρικές μεθόδους (vector spaces, graphs).

References:

- [1] D. Widdows, “Geometry and Meaning”, Vol. 773. Stanford: CSLI publications, 2004.
- [2] P. Gärdenfors, “Conceptual spaces: The Geometry of Thought”, MIT Press, 2004.
- [3] T. Mikolov et al. [“Efficient estimation of word representations in vector space”](#), ICLR 2013.