Proposition d'atelier pour PFIA2026

Contributeurs : Faten Chaieb-Chakchouk et Djallel DILMI

Titre proposé :

Des modèles de langage aux modèles visuels : mécanismes d'attention (Vaswani et al. 2017) et applications en vision par ordinateur

Description courte (2 phrases)

Ce tutoriel explore les mécanismes d'attention, depuis leur émergence dans les modèles de langage (LLM) jusqu'à leur adaptation en vision par ordinateur. Il met l'accent sur les fondements mathématiques, les intuitions conceptuelles et les applications pratiques en vision.

Description longue (2 paragraphes)

Les mécanismes d'attention ont profondément transformé l'intelligence artificielle moderne, en particulier avec les modèles de langage de grande taille (LLM) basés sur les Transformers. Ce tutoriel présente une introduction progressive aux concepts fondamentaux de l'attention, non pas comme une simple opération calculatoire, mais comme un produit scalaire dynamique entre représentations, avec un lien direct avec des méthodes classiques telles que l'analyse canonique des corrélations (CCA). Cette approche permet de comprendre pourquoi l'attention capture efficacement les dépendances complexes et les relations contextuelles.

Dans une seconde partie, le tutoriel se concentre sur l'adaptation de ces mécanismes au domaine de la vision par ordinateur. Seront abordés les Vision Transformers -ViT- (Dosovitskiy et al. 2020), l'attention spatiale et canal -SE, CBAM, Mamba- (Wang et al. 2024, Gu et al. 2023), ainsi que les architectures hybrides CNN + attention. Une session pratique permettra aux participants de manipuler des modèles d'attention appliqués à des données visuelles, de visualiser les cartes d'attention et de comprendre leurs avantages et limites dans des tâches de classification et de détection.

Déroulé détaillé (3h)

Étape / Durée	Contenu
Introduction (15–20 min)	• Évolution des architectures : CNN → Transformers → LLM • Pourquoi l'attention ? Intuition : "où regarder" dans les données • Objectifs pédagogiques
Fondements mathématiques des mécanismes d'attention (40 min)	Objectif pédagogique : comprendre l'origine et la rationalité mathématique de l'attention, au-delà de la simple définition calculatoire. • Lecture mathématique : attention comme produit scalaire dynamique

	$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$ mais interprété comme mesure de similarité adaptative entre représentations • Produit scalaire dynamique : $$\langle x,y\rangle_A = x^T Ay$$ A dépendant du contexte (appris) • Lien avec l'analyse canonique des corrélations (CCA) : o CCA cherche des projections u, v maximisant $\text{corr}(Xu, Yv)$ o Attention : projections apprises $Q = XW_Q, K = YW_K$, similarité calculée localement **Message clé : généralisation de méthodes statistiques classiques pour créer un produit scalaire contextuel et dynamique**
Attention dans les modèles de langage (LLM) (30 min)	• Self-attention et multi-head attention • Architecture Transformer simplifiée • Compréhension de la dépendance globale dans les séquences
Passage du NLP à la vision (30 min)	• Différences texte vs image : structure spatiale vs séquentielle • Tokenisation des images (patchs) • Adaptation des mécanismes d'attention à l'espace visuel
Attention en vision par ordinateur (30 min)	• Vision Transformers (ViT) • Attention spatiale et canal (SE, CBAM) • Architectures hybrides CNN + attention
Session pratique et discussion (60 min)	• Implémentation d'un module de self-attention (TensorFlow ou PyTorch) • Application sur des datasets visuels (CIFAR-10, subset ImageNet et des données médicales) • Visualisation et interprétation des cartes d'attention Discussion : • Limites : coût computationnel, robustesse et interprétabilité • Extensions récentes et applications multimodales • Ouverture vers attention contextuelle adaptative

Public cible

- Doctorants, chercheurs et ingénieurs en IA
- Enseignants souhaitant introduire l'attention dans leurs cours
- Niveau : intermédiaire

Références

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit and Neil Houlsby. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” *ArXiv* abs/2010.11929 (2020): n. pag.

Gu, Albert and Tri Dao. “Mamba: Linear-Time Sequence Modeling with Selective State Spaces.” *ArXiv* abs/2312.00752 (2023): n. pag.

Vaswani Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

Wang, Y., Wang, W., Li, Y. et al. An attention mechanism module with spatial perception and channel information interaction. Complex Intell. Syst. 10, 5427–5444 (2024). https://doi.org/10.1007/s40747-024-01445-9