



Sign Spotting dans les vidéos de Langue de Signes Française pour l'enseignement des mathématiques

Annelies Braffort, Claire Dartyge, Michèle Gouiffès

Contacts : michele.gouiffes@lisn.upsaclay.fr, annelies.braffort@lisn.upsaclay.fr, claire.dartyge@math.univ-toulouse.fr

Mots clés : Vision par ordinateur, vidéos, apprentissage automatique, langue des signes française, enseignement des mathématiques.

Dates du stage : 5 mois, approximativement de mars à juillet 2023

Gratification : 591,51€ mensuels + remboursement de frais de transport (50% abonnement mensuel ou annuel Navigo).

Lieu du stage : Laboratoire Interdisciplinaire des Sciences du Numérique ([LISN](#)) situé à Orsay, Campus universitaire bât 507, Rue du Belvédère.

1 Contexte

Les Langues des Signes (LS) sont des langues naturelles pratiquées au sein des communautés de Sourds, avec leur lexique (les « signes ») et grammaire propre qui est différente de celle des langues parlées : la grammaire est spatiale et très iconique (on peut montrer de ce l'on dit). Ce sont des langues visuo-gestuelles multilinéaires qui utilisent les mains mais aussi la posture du buste et des épaules, les expressions du visage et la direction du regard. Il n'y a pas une langue des signes universelle puisque ce sont des langues naturelles. Elles sont aussi variées que les langues parlées. En France, on pratique la Langue des Signes Française ou LSF.

La recherche sur le traitement automatique des langues des signes est très récente comparée à celle sur les langues parlées ou écrites [[Bragg et al., 2019](#)]. À ce jour, les LS n'ont pas de système d'écriture et elles sont la plupart du temps sous forme de vidéos. Ces vidéos sont de natures très diverses : interprétation en direct à la télévision, média en ligne bilingue (LSF avec sous-titrage en français), ressources pédagogiques en ligne tel que le site [Sign'Maths](#).

Ces langues sont très peu outillées (pas d'outils tels qu'il en existe pour les langues parlées ou écrites) et la demande sociétale est pressante, en particulier pour l'outillage des vidéos. C'est dans ce contexte que se situe le stage.

2 Sujet

Sign'Math est un site web à visée pédagogique dédié à l'enseignement des mathématiques en LSF. Il comporte un glossaire bilingue LSF/français sur les notions mathématiques pour l'enseignement en LSF de la primaire à l'enseignement supérieur. Les signes disponibles sur Sign'Maths sont soit des signes déjà répertoriés dans des dictionnaires de LSF, soit des néologismes proposés au sein d'un groupe de travail réunissant des enseignants de mathématiques en LSF sourds et entendants de plusieurs régions de France.

C'est donc un lexique en évolution constante mais de taille limitée. Certains de ces signes sont construits par combinaison de plusieurs signes plus élémentaires. La plupart d'entre eux sont fortement illustratifs ou iconiques.

L'objectif du stage est d'étudier un moteur de recherche visuel qui permettrait, à partir d'une vidéo désignant une notion de mathématiques en LSF, d'obtenir une liste de vidéos qui contiennent cette notion. Par exemple, si on choisit la notion de « suite », la recherche devra retourner toutes les vidéos qui comportent le signe correspondant à cette notion (par exemple, suite croissante, suite monotone, suite convergente, etc.).

Les travaux qui seront entamés pendant le stage visent à terme à contribuer au développement du site web Sign'Maths, en fournissant un moteur de recherche par vidéo. Le site dispose déjà d'un moteur de recherche par le texte.

Nous travaillerons plus particulièrement sur les similarités gestuelles. Étant donnée la taille des données disponibles, nous nous pencherons sur les approches de type *zero-shot* [Bilge et al., 2019], *few shot* spotting [Dhillon et al., 2020, Momeni et al., 2020] et les approches de *Contrastive learning*, que nous adapterons à nos objectifs, en utilisant éventuellement du transfert depuis des corpus plus conséquents mais non spécialisés au domaine des mathématiques.

Pour ce faire, nous exploiterons les travaux réalisés précédemment sur la reconnaissance de signes dans les vidéos de LSF [Belissen et al., 2020]. Les vidéos d'entrée seront encodées soit par des modèles convolutionnels spatio-temporels [Carreira and Zisserman, 2017] et/ou détecteurs de pose humaine par points clés [Dafnis, 2022], pré-entraînés sur des bases plus conséquentes.

3 Encadrement

Le stage se déroulera au sein des équipes AMI et M&TALS du LISN, qui travaillent respectivement sur la vision par ordinateur pour l'interaction et les langues des signes (corpus, modèles, reconnaissance et synthèse), avec une collaboration avec l'équipe de Sign'Maths.

4 Compétences requises

Nous recherchons une étudiante ou un étudiant en master 2 intéressé par le monde de la recherche et ayant une formation solide en mathématiques et informatique avec une spécialisation en vision par ordinateur et machine learning. Le candidat développera en Python (Tensorflow/ pytorch). La connaissance de la LSF n'est pas exigée.

Références

- [Belissen et al., 2020] Belissen, V., Braffort, A., and Gouiffès, M. (2020). Experimenting the automatic recognition of non-conventionalized units in sign language. *Algorithms*, 13(12) :310.
- [Bilge et al., 2019] Bilge, Y. C., Ikizler-Cinbis, N., and Cinbis, R. G. (2019). Zero-shot sign language recognition : Can textual data uncover sign languages ?
- [Bragg et al., 2019] Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., and ... Vogler, C. (2019). Sign language recognition, generation, and translation : An interdisciplinary perspective. In *21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 16–31.
- [Carreira and Zisserman, 2017] Carreira, J. and Zisserman, A. (2017). Quo vadis, action recognition ? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733.
- [Dafnis, 2022] Dafnis, Konstantinos M. ; Chroni, E. N. C. M. D. (2022). Bidirectional skeleton-based isolated sign recognition using graph convolution networks. *LREC 2022*.
- [Dhillon et al., 2020] Dhillon, G. S., Chaudhari, P., Ravichandran, A., and Soatto, S. (2020). A baseline for few-shot image classification. In *International Conference on Learning Representations*.
- [Momeni et al., 2020] Momeni, L., Varol, G., Albanie, S., Afouras, T., and Zisserman, A. (2020). Watch, read and lookup : learning to spot signs from multiple supervisors. *arXiv preprint*, (2010.04002).