**Ressources nécessaires**

Présentation du projet et des principaux concepts (issus des slides de présentation) :

Segmentation médicale : gérer l’incertitude et la variabilité inter-experts.

Description : Segmenter les organes est parfois très complexe car les organes varient beaucoup en position et forme et que plusieurs experts ne sont pas toujours d’accord pour les segmentations de radiographies. Le but du projet est de garantir de meilleurs diagnostiques en créant un algorithme capable de gérer ces incertitudes. 2 types d’incertitude : aléatoire (données) et épistémique (modèle). But de la quantification des incertitudes : 1) carte d’incertitude pour identifier les zones peu sûre, 2) fournir des prédictions calibrées pour les cliniciens, 3) sécurité accrue en intégrant l’incertitude dans les décisions médicales.

Une image contenant texte, capture d’écran, diagramme, Tracé

Description générée automatiquement

Étapes :

1. Familiarisation avec le contexte et les outils.
2. Reproduction de la baseline (entrainement de nnU-Net et analyse des métriques Dice – ECE – CRPS).
3. Analyse de l’incertitude via ValUES.

Données et code :

* **Pour Statapp :** [**https://github.com/Kirscher/statapp\_2025\_curvas**](https://github.com/Kirscher/statapp_2025_curvas)
* **CURVAS :** [**https://curvas.grand-challenge.org/curvas/**](https://curvas.grand-challenge.org/curvas/)
  + Calibration and Uncertainty for multiRater Volume Assessment in multiorgan Segmentation. Addressing interrater variability in DL for medical segmentation involves the development of robust algorithms capable of capturing and quantifying uncertainty, as well as standardizing annotation practices and promoting collaboration among medical experts to reduce variability and improve the reliability of DL-based medical image analysis. The images are CT scans.
* **Repository CURVAS:** [**https://github.com/SYCAI-Technologies/curvas-challenge**](https://github.com/SYCAI-Technologies/curvas-challenge)
  + Contains the training set.
* **Repository ValUES:** [**https://github.com/IML-DKFZ/values**](https://github.com/IML-DKFZ/values)
  + ValUES is a framework for systematic validation of uncertainty estimation in semantic segmentation. Can **data-related (aleatoric)** and **model-related (epistemic)** uncertainty really be separated in practice? Which components of an uncertainty method are essential for real-world performance? Which uncertainty method works well for which application? In this work, we link this research gap to a lack of systematic and comprehensive evaluation of uncertainty methods. Specifically, we identify three key pitfalls in current literature and present an evaluation framework that bridges the research gap by providing 1) a controlled environment for studying data ambiguities as well as distribution shifts, 2) systematic ablations of relevant method components, and 3) test-beds for the five predominant uncertainty applications: OoD-detection, active learning, failure detection, calibration, and ambiguity modeling.

Une image contenant texte, capture d’écran, Police, logiciel

Description générée automatiquement

* **Repositoy nnU-Net:** [**https://github.com/MIC-DKFZ/nnUNet**](https://github.com/MIC-DKFZ/nnUNet)
  + nnU-Net is a deep learning algorithm for semantic segmentation that automatically adapts to a given image dataset (2D, 3D, RGB image, CT, MRI, …). It will analyze the provided training cases and automatically configure a matching U-Net-based segmentation pipeline. No expertise is required. You can simply train the models and use them. nnU-net relies on supervised learning and uses a lot of data augmentation.

Articles à lire :

* Prioritaires :

Isensee, Fabian, et al. nnU-Net : Self-Adapting Framework for U-Net-Based Medical Image Segmentation. arXiv, 2018. **11 pages** <https://arxiv.org/abs/1809.10486>

Lambert, Benjamin, 2024, Quantification et caractérisation de l’incertitude de segmentation d’images médicales par des réseaux profonds, PhD thesis, Université Grenoble Alpes, https://theses.hal.science/tel-04673383 **328 pages** [**https://theses.hal.science/tel-04673383**](https://theses.hal.science/tel-04673383)

Kahl, Kim-Celine, et al. ValUES : A Framework for Systematic Validation of Uncertainty Estimation in Semantic Segmentation. arXiv, 2024. **35 pages** [**https://arxiv.org/abs/2401.08501**](https://arxiv.org/abs/2401.08501)

Riera i Marín, Meritxell, et al. CURVAS : Calibration and Uncertainty for MultiRater Volume Assessment in Multiorgan Segmentation. Apr. 2024. **21 pages** <https://zenodo.org/records/10979642>

Maier-Hein, Lena, Matthias Eisenmann, Annika Reinke, et al., 2018, Why rankings of biomedical image analysis competitions shouldbe interpreted with care, Nature Communications, 9:5217 **13 pages** <https://www.nature.com/articles/s41467-018-07619-7>

Maier-Hein, Lena, Annika Reinke, et al. Metrics Reloaded : Recommendations for Image Analysis Validation. Nature Methods, vol. 21, no. 2, Feb. 2024, pp. 195–212. **30 pages** <https://www.nature.com/articles/s41592-023-02151-z>

Ronneberger, Olaf, et al. U-Net : Convolutional Networks for Biomedical Image Segmentation. arXiv, 2015. **8 pages** [**https://arxiv.org/abs/1505.04597**](https://arxiv.org/abs/1505.04597)

* En plus :

Learning to count biological structures with raters' uncertainty <https://zenodo.org/records/8326453>

Multi-rater Prompting for Ambiguous Medical Image Segmentation <https://arxiv.org/pdf/2404.07580>

Issu du site web CURVAS :

Exploitation de l’incertitude pour améliorer les modèles :

* **Out-of-Distribution Detection (OoD)** : Détecter les cas hors distribution pour éviter les erreurs.
* **Failure Detection (FD)** : Identifier les échecs de segmentation et détecter les prédictions erronées.
* **Active Learning (AL)** : Sélectionner les échantillons incertains à annoter pour améliorer l’apprentissage.
* **Calibration** : Ajuster les prédictions du modèles pour qu’elles reflètent correctement les probabilités d’incertitude.
* **Ambiguity Modelling (AM)** : Modéliser l’incertitude pour quantifier la confiance des prédictions.

Images regions:

* **Foreground consensus areas**: Regions on which the three clinicians have agreed that belongs to the foreground.
* **Background consensus areas**: Regions on which the three clinicians have agreed that belongs to the background.
* **Dissensus areas**: Regions in which the three labelers did not agree.

Metrics:

* **Quality of the segmentation and uncertainty consensus**:
  + **Dice score** **(DSC)**: assesses only the consensus foreground and background areas. Any prediction within the dissensus area will be ignored. False Positives can only occur in the consensus background, False Negatives can only occur in the consensus foreground area. The mean confidence for the consensus background and foreground are **CB**and **CF**. The confidence assessment performed in the consensus is **Cseg = [(1 – CB) + CF]/2**. The mean of the three confidence values obtained per each class will be extracted to determine the overall confidence metric for evaluating uncertainty.

Une image contenant texte, capture d’écran, Police, document

Description générée automatiquement

* **Multi-rater calibration**:
  + **Expected Calibration Error** **(ECE)**: To maintain the multirater information, calibration will be computed for each prediction against each annotation, resulting in three ECE values. We then take the average.

Une image contenant texte, Police, blanc, capture d’écran

Description générée automatiquement

* **Volume assessment**:
  + The metrics designed above are not relevant for real-world scenarios. To address this, we explore the study of biomarkers such as volume. For this volume assessment, we adopted a different approach. First, we needed a method to retain information from the various annotations. Thus, we defined a gaussian Probabilistic Distribution Function (PDF) using the mean and standard deviation derived from computing the volume the three different annotations. Once the PDF was established, we then defined the corresponding Cumulative Distribution Function (CDF). The predicted volume will be calculated by summing all the probabilistic values for the corresponding class from the probabilistic matrix provided by the participant. This method considers the model's uncertainty and confidence in its predictions when evaluating the volume. For the evaluation of this section the **Continuous Ranked Probability Score** (CRPS) will be used, see Equation 3. The CRPS measures the average squared difference between a cumulative distribution and a predicted value.

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement

Une image contenant texte, ligne, Tracé, diagramme

Description générée automatiquement

Tâche pour aujourd’hui : regarder le git hub, créer ma branche (mais demander à chat gpt comment le faire de manière professionnelle) mettre mon document word à continuer de compléter, regarder quels articles j’ai besoin de lire pour faire la première tâche (entrainer le modèle avec les données de CURVAS).