# Active learning for medical image segmentation with stochastic batches

**Mélanie Gaillochet**[1] melanie.gaillochet.1@ens.etsmtl.ca
**Christian Desrosiers**[1] christian.desrosiers@etsmtl.ca
**Hervé Lombaert**[1] herve.lombaert@etsmtl.ca [1] *ETS Montréal*

## Abstract

Active learning (AL) selects informative samples for annotation. This is becoming increasingly crucial to medical image segmentation since image annotation is hardly scalable to full pixel-level labeling of large datasets. However, most research focuses on classification or natural image segmentation. Uncertainty-based AL methods tend to have sub-optimal batch-query strategies, and diversity-based methods are computationally expensive. This work improves uncertainty-based AL for medical image segmentation using stochastic batches during sampling, computing uncertainty at the batch-level. Experiments on MRI prostate imaging show this approach's effectiveness and robustness under various conditions.

**Keywords:** Active learning, Segmentation, Medical image analysis, Uncertainty.

## 1. Introduction

The performance of deep learning-based segmentation algorithms relies on annotated training data. However, manual annotation is laborious and costly, particularly in the medical domain. To solve this constraint, active learning (AL) (Settles, 2009) identifies the most valuable samples to annotate, maximizing model performance with minimal labelled data.

AL methods applied to deep learning-based models (deep AL) include uncertainty-based, representative-based, and hybrid approaches. Uncertainty-based strategies (Wang et al., 2017; Beluch et al., 2018; Kirsch et al., 2019; Yoo and Kweon, 2019) target samples for which the current model is least confident, but may focus on outliers. Representative-based strategies (Sener and Savarese, 2018; Sinha et al., 2019) and hybrid AL approaches (Ash et al., 2020; Kim et al., 2021; Nath et al., 2021) ensure diversity but struggle with high dimensionality, often limiting their application to classification tasks. Furthermore, existing work on pixel-wise annotations primarily targets natural image segmentation. To our knowledge, few deep AL strategies have explored medical image segmentation, and they they are often computationally expensive and difficult to scale (Yang et al., 2017; Nath et al., 2021). There hence remains a gap between active learning and medical image segmentation.

An increasing number of AL studies recognize the difficulty in outperforming random sampling (Mittal et al., 2019). Gains from AL strategies over random sampling are often inconsistent across various experimental setups. In addition, the sensitivity of existing AL methods to factors such as training hyperparameters or regularization often means that observed improvements can be easily voided (Munjal et al., 2022). These challenges increase the difficulaty of applying AL to medical image segmentation.

This paper tackles AL limitations by incorporating randomness in uncertainty-based batch sampling to improve medical imaging segmentation. We propose to use stochastic batch (SB) querying alongside existing uncertainty-based AL strategies (see Figure 1). Our novel approach provides several benefits:

1. a sampling strategy that can tackle **medical image segmentation**;
2. a **flexible framework** compatible with any uncertainty-based AL strategy; and
3. a **scalable** method ensuring diverse sample selection.

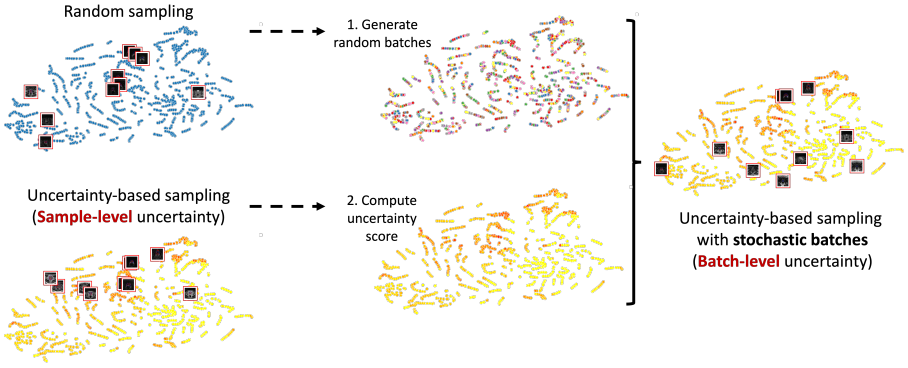



Figure 1: **Stochastic batch AL for uncertainty-based sampling**. The diversity from random sampling is combined with the informativeness of uncertainty-based sampling.

## 2. Method

We initially train a segmentation model $f_\theta(\cdot)$ on an randomly sampled labeled set $\mathcal{D}_L = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$. After the first training cycle, we select $B$ candidate samples from the unlabeled set $\mathcal{D}_U = \{x_u^{(j)}\}_{j=1}^{M}$, which we annotate and add to the labeled training set $\mathcal{D}_L$. This process is repeated until the annotation budget is exhausted.

Our AL method uses stochastic batches in two stages for guided sampling diversity. First we generate $Q$ batches, each with $B$ samples selected uniformly at random from $\mathcal{D}_U$:

$$Batch^{(i)} = \{x_u^{(i_1)}, x_u^{(i_2)}, ..., x_u^{(i_B)}\} \sim Uniform(\mathcal{D}_u, B).$$

Then, for each generated batch, we assign an uncertainty score to each unlabeled sample it contains, according to the current model $f_{\hat{\theta}}$ and the chosen uncertainty metric ($m_{unc}$), and we compute the mean $u_{score}$ for the batch:

$$\forall k = 1, ..., B: \quad u_{score}^{x_u^{(i_k)}} = m_{unc}(f_{\hat{\theta}}, x_u^{(i_k)}),$$

$$u_{score}^{Batch^{(i)}} = \frac{1}{B} \sum_{k=1}^{B} u_{score}^{x_u^{(i_k)}}.$$

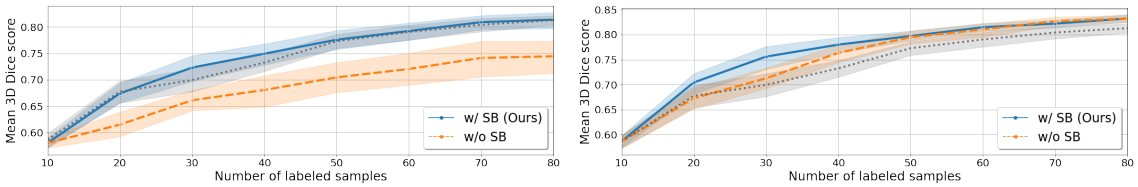
The batch with the highest mean score yields the annotation candidates.

## 3. Results

We validate our method on the PROMISE 2012 dataset (Litjens et al., 2014), with 248 test images from 10 patients. Initially, $|\mathcal{D}_L| = 10$, $|\mathcal{D}_U| = 1010$. After each AL cycle, $B = 10$ new samples are selected via AL, annotated and added to $\mathcal{D}_L$. The model, a UNet (Ronneberger et al., 2015), is then retrained from scratch. Each experiment is repeated 5 times.

Table 1: **Improvements with Stochastic Batches over varying initial labelled sets.** Mean model performance over all AL cycles. Adding stochastic batches provides an improvement at a statistically significant level (indicated by *).

| | RS | Entropy | | Dropout | |
|---|---|---|---|---|---|
| | | w/o SB | w/ SB (ours) | w/o SB | w/ SB (ours) |
| **3D DSC** | 68.8 (±16.0) | 67.0 (±16.7) | **71.3*** (±17.4) | 67.7 (±17.2) | **72.6*** (±15.0) |
| **2D DSC** | 67.9 (±8.3) | 66.9 (±8.6) | **69.0*** (±9.0) | 67.1 (±9.5) | **69.6*** (±8.1) |
| **3D HD95** | 7.0 (±3.7) | 7.0 (±4.3) | **6.7*** (±3.1) | 7.0 (±5.0) | **6.6*** (±3.2) |




(a) Improvement for Learning loss (Yoo and Kweon, 2019) (b) Improvement for TTA (Gaillochet et al., 2022)

Figure 2: **Improvements with Stochastic Batches over varying hyperparameters**. Stochastic batches improve the model performance of purely uncertainty-based AL strategies, regardless of the initial labelled set.

We set $Q = \frac{|\mathcal{D}_U|}{B}$ and compare our stochastic batch strategy with random sampling (RS) and four purely uncertainty-based methods: Entropy-based sampling (Shannon, 1948), Dropout-based sampling (Gal and Ghahramani, 2016), Learning Loss (Yoo and Kweon, 2019) and TTA-based sampling (Gaillochet et al., 2022).

Table 1 depicts the average results over 5 different initial labelled sets, across all AL cycles. Our stochastic batch selection strategy improves purely uncertainty-based selection at a statistcally significant level even when we vary the initial labelled set. Figure 2 shows the mean model performance over trainings with different hyperparameters, across AL cycles. Adopting stochastic batches during sampling yields a significant boost in terms of 3D dice score. This jump in performance (orange to blue) is particularly notable during the first four AL cycles, reaching above the random sampling baseline (dotted grey).

## 4. Conclusion

Active learning is becoming increasingly crucial in medical image segmentation since annotating full large datasets may be unrealistic with clinical time constraints. This paper tackles three key limitations of AL strategies: the scarcity of AL work in medical image segmentation, the tendency of uncertainty-based batch sampling to select similar samples, and the computational burden of diversity-based methods. Our approach computes uncertainty at the batch level with randomly generated sample batches, offering a simple, computationally-efficient way to improve AL candidate selection and model performance. Our method proves effective for the complex task of medical image segmentation, improving uncertainty-based AL strategy while being robust to variations in training settings.

## References

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. In *Eighth International Conference on Learning Representations (ICLR)*, 2020.

W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler. The Power of Ensembles for Active Learning in Image Classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9368–9377, 2018.

Mélanie Gaillochet, Christian Desrosiers, and Hervé Lombaert. Taal: Test-time augmentation for active learning in medical image segmentation. In *Data Augmentation, Labeling, and Imperfections (MICCAI DALI)*, LNCS, pages 43–53, 2022.

Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML)*, volume 48 of *ICML'16*, pages 1050–1059, 2016.

Kwanyoung Kim, Dongwon Park, Kwang In Kim, and Se Young Chun. Task-Aware Variational Adversarial Active Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8162–8171, 2021.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, Robin Strand, Filip Malmberg, Yangming Ou, Christos Davatzikos, Matthias Kirschner, Florian Jung, Jing Yuan, Wu Qiu, Qinquan Gao, Philip ËddieËdwards, Bianca Maan, Ferdinand van der Heijden, Soumya Ghose, Jhimli Mitra, Jason Dowling, Dean Barratt, Henkjan Huisman, and Anant Madabhushi. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Medical Image Analysis*, 18(2):359–373, 2014.

Sudhanshu Mittal, Maxim Tatarchenko, Ozgün Çiçek, and Thomas Brox. Parting with Illusions about Deep Active Learning. *arXiv:1912.05361*, 2019.

Prateek Munjal, Nasir Hayat, Munawar Hayat, Jamshid Sourati, and Shadab Khan. Towards Robust and Reproducible Active Learning Using Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Vishwesh Nath, Dong Yang, Bennett A. Landman, Daguang Xu, and Holger R. Roth. Diminishing Uncertainty Within the Training Pool: Active Learning for Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 40(10):2534–2547, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, pages 234–241. Springer, 2015.

Ozan Sener and Silvio Savarese. Active Learning for Convolutional Neural Networks: A Core-Set Approach. In *International Conference on Learning Representations (ICLR)*, 2018.

Burr Settles. Active Learning Literature Survey. Technical Report, University of Wisconsin-Madison Department of Computer Sciences, 2009.

C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Samrath Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational Adversarial Active Learning. In *IEEE International Conference on Computer Vision (ICCV)*, pages 5971–5980, Seoul, Korea (South), 2019.

Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-Effective Active Learning for Deep Image Classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27:2591–2600, 2017.

Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z. Chen. Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 10435 of *Lecture Notes in Computer Science*, pages 399–407. Springer, 2017.

Donggeun Yoo and In So Kweon. Learning Loss for Active Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 93–102, 2019.