Deep Image Segmentation based on Mutual Information: A Study

Tiago Gonçalves, Leonardo Capozzi, Ana Rebelo, Jaime S. Cardoso

Introduction

Image segmentation is a computer vision problem where the goal is to classify individual pixels of an image. Information theory is based on probability and statistics and is often used to study the information of the distributions of random variables. Information theory concepts such as the Kullback-Leibler Divergence (KLD), are widely used as optimisation criteria

In this paper, we present a comparative study of two methodologies based on mutual information applied to the task of deep image segmentation

Information Theory Concepts

Information

$$I_X(x) = -log_2[p_X(x)]$$

Mutual Information

$$I(X;Y) = \sum_{x,y} p(x,y) log_2 \frac{p(x,y)}{p(x)p(y)}$$

Kullback-Leibler Divergence

$$D_{KL}(P \parallel Q) = \sum_{x \in X} P(x) log(\frac{P(x)}{Q(x)})$$

Data

We performed experiments with two benchmark data sets: **1) the Cityscapes Dataset**, composed of urban street scenes recorded from 50 different cities; and, **2) the PASCAL Visual Object Classes (VOC) 2012**, composed of several visual object classes in realistic scenes

This data processing strategy is employed in all the use-cases: 1) images and masks are read and are given to the models in RGB format; 2) the images are then resized into the size 234 × 234; 3) a random-crop is then applied to the images and they reach the final size of 224×224

Models

We trained three different models: **1) U-Net Baseline (Model 1)**; **2) U-Net with Region Mutual Information Loss (Model 2)**; and, **3) Deep Mutual Learning using DeepLabV3 and U-Net (Model 3)**

In Model 2, we train a U-Net with the loss proposed by [3], given by **L=λLBCE+(1−λ)LRMI**, where **LBCE** is the binary cross-entropy loss, **LRMI** is the region mutual information loss and λ=0.5 is a weight coefficient applied to the loss terms

In Model 3, we apply the training strategy proposed by [2], where we employ a joint training strategy with a teacher model and a student model. In this work, we used the DeepLabV3 architecture as the teacher model and U-Net as the student model

Performance Metrics

To assess the predictive performance of the deep learning algorithms we use two performance metrics: **1) intersection over union (IoU)** and **2) accuracy (Acc)**, defined as:

$$\text{IoU} = \frac{TP}{TP+FP+FN}$$

$$\text{Acc} = \frac{TP+TN}{TP+FP+FN+TN}$$

, where **TP**, **TN**, **FP** and **FN** be the number of true positive pixels, the number of true negative pixels, the number of false-positive pixels and the number of false-negative pixels, respectively

Results, Discussion and Conclusions

Table 1 presents the IoU and Acc values obtained for Model 1, Model 2 and Model 3

Despite the clear differences between the two datasets, we note that predictive performance does not significantly change for Model 2, and Model 3

Figures 1–2 show examples of predictions on the PASCAL VOC 2012 dataset

Table 1: Summary of results (IoU and Acc) for all models and datasets. Best results highlighted in bold.

Dataset	Model 1		Model 2		Model 3	
	IoU	Acc	IoU	Acc	IoU	Acc
Cityscapes Dataset	**0.5931**	**0.9388**	0.5244	0.9345	0.5619	0.9342
PASCAL VOC 2012	0.1965	0.7367	0.0607	0.6893	**0.2203**	**0.7539**

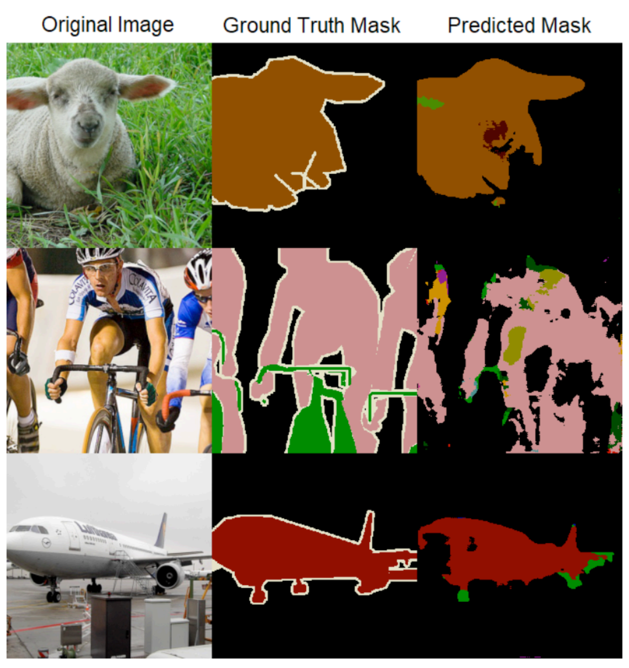

Figure 1: Examples of predicted masks for the PASCAL VOC 2012 using Model 2.

Figure 2: Examples of predicted masks for the PASCAL VOC 2012 using Model 3.

The application of the Region Mutual Information Loss does not seem to bring improvements to the predictive performance of the U-Net model in these data sets (some authors suggest that some of the reported results regarding the benefits of using mutual information to optimise deep learning algorithms, may not be attributed to the properties of mutual information alone [1])

Further work should be devoted to the evaluation of different deep image segmentation architectures on the same use-case and to the study of different mutual information properties that could be incorporated in the regularisation of different deep learning algorithms to assess if these methodologies are data set

References

[1] Michael Tschannen et al. On mutual information maximization for representation learning. arXiv preprint arXiv:1907.13625, 2019.
[2] Ying Zhang et al. Deep mutual learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4320–4328, 2018.
[3] Shuai Zhao et al. Region mutual information loss for semantic segmentation. arXiv preprint arXiv:1910.12037, 2019.

Acknowledgements

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project NewSpacePortugal, with reference 11, and by FCT – Fundação para a Ciência e a Tecnologia within he PhD grants “2020.06434.BD” and “2021.06945.BD”.