# Uma abordagem Bayesiana na tarefa de segmentação semântica

**João Lucas Foltran Consoni**
**Nícolas dos Santos Rosa**
**Prof. Dr. Marco Henrique Terra**
EESC/USP
joaofconsoni@usp.br, nicolas.rosa@usp.br, terra@sc.usp.br

## 1 Objetivos

Este trabalho tem como objetivo principal estimar incertezas aleatórias e epistêmicas para redes neurais profundas utilizando métodos de inferência Bayesiana, com enfoque na tarefa de segmentação semântica para veículos autônomos operando em cenário urbano e rodoviário. Mais especificamente, propôs-se modificar uma arquitetura para a tarefa mencionada que não utiliza a abordagem Bayesiana. Desse modo, ao incorporar a estimativa de incertezas na rede, busca-se identificar as regiões confiáveis e as de maior incerteza na predição.

## 2 Métodos e Procedimentos

### 2.1 Inferência Bayesiana

A rede neural deve ser modelada com conceitos de inferência Bayesiana, de modo que possa estimar as incertezas associadas às predições. Os pesos da rede são inicializados por alguma distribuição, denotada por $p(\mathbf{W})$, como uma Gaussiana, por exemplo. Denotando a saída da rede Bayesiana por $\mathbf{f}^{\mathbf{W}}(\mathbf{x})$, para uma entrada $\mathbf{x}$ qualquer, a função de probabilidade do modelo é definida por $p(\mathbf{Y}|\mathbf{X}, \mathbf{W})$. (KENDALL; GAL, 2017). Para problemas de classificação, a probabilidade apresentada acima pode ser descrita por:

$$p(\mathbf{y}|\mathbf{x}, \mathbf{W}) = \text{Softmax}(\mathbf{f}^{\mathbf{W}}(\mathbf{x})). \quad (1)$$

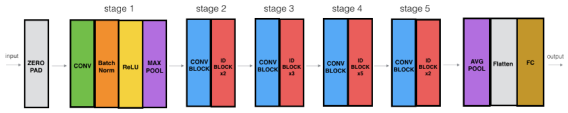
Para capturar a incerteza aleatória associada com a rede neural, deve-se maximizar a função $p(\mathbf{Y}|\mathbf{X}, \mathbf{W})$. Seja $\hat{\mathbf{W}}_{MLE}$ o estimador de máxima verossimilhança para os parâmetros de $\mathbf{W}$. $\hat{\mathbf{W}}_{MLE}$ pode ser determinado minimizando $-\sum_{i=1}^{N} \log p(\mathbf{y_i}|\mathbf{x_i}, \mathbf{W})$, que é a função de custo NLL (*Negative Log-Likelihood*).

Para capturar a incerteza epistêmica da rede neural, tem-se como objeto de interesse a distribuição $p(\mathbf{W}|\mathbf{X}, \mathbf{Y})$, denominado *posterior*. Para grandes redes neurais, computacionalmente, a determinação dessa distribuição é difícil (KENDALL; GAL, 2017).

### 2.2 Arquitetura da rede

A rede utilizada no projeto consiste em uma *ResNet50* como base (*Trunk*) e um *decoder* dedicado a segmentação semântica (*Segmentation Head*), semelhante a que foi utilizada por (TAO; SAPRA; CATANZARO, 2020). A arquitetura mencionada pode ser vista na Figura 1.

Figura 1: Arquitetura da ResNet50




Fonte: Imagem retirada de (DWIVEDI, 2019).

### 2.3 Desenvolvimento do projeto

O projeto foi desenvolvido na linguagem de programação *Python*, utilizando a API *keras* (CHOLLET et al., 2015), e a rede foi treinada com uma GPU NVIDIA® Titan Xp (12GB). Tanto para o treinamento e para avaliação da rede utilizou-se o *dataset Cityscapes* (CORDTS et al., 2016).

O modelo, inicialmente, possuia uma *Segmentation Head* semelhante a utilizada por Tao, Sapra, Cantazaro (2020). Primeiramente, foi adicionada uma camada *OneHotCategorical*, da biblioteca de camadas probabilísticas do *tensorflow*, após a convolução 1x1, para estimar a incerteza aleatória da rede. Depois, as camadas convolucionais foram substituídas por camadas *Convolution2DReparameterization* para estimar a incerteza epistêmica da rede, que não convergiu nesse caso.

Portal de Eventos Científicos da Escola de Engenharia de São Carlos
Anais eletrônicos do 30º Simpósio Internacional de Iniciação Científica e Tecnológica da USP & 36º Congresso de Iniciação Científica e Tecnológica em Engenharia
Disponível em: http://eventos.eesc.usp.br/siicusp2022
Serviço de Biblioteca Prof. Dr. Sérgio Rodrigues Fontes

Para ambos os treinamentos, a escala utilizada foi 1024x2048, o optimizador foi o SGD, com *learning rate* inicial de 0,001 e *batch size* de 2 imagens. Como *data augmentation* foi utilizada a transformação *color jitter*.

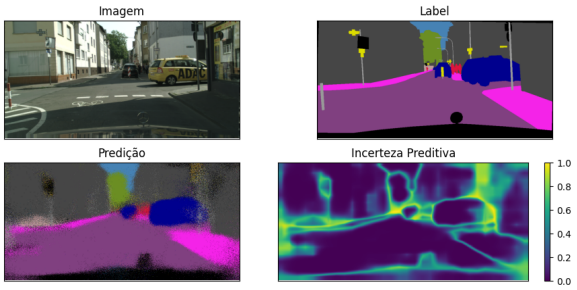
## 3 Resultados

Os resultados apresentados são da rede que estima apenas a incerteza aleatória da rede, visto que a outra rede não convergiu. Para o treinamento da rede foram utilizadas 19 classes do *dataset*, como em (TAO; SAPRA; CATANZARO, 2020), mais uma classe de *background*.

Para a avaliação do modelo, foram utilizados os dados de validação do *dataset*. Como validação, foi utilizada uma porção do *dataset* de treino. A métrica utilizada foi a mIoU (*mean Intersection over Union*), comumente usada para mensurar a qualidade da predição desta tarefa.

A predição e incerteza da primeira imagem usada para teste é mostrada na Figura 2.

Figura 2: Predição e incerteza associadas




Fonte: Imagem elaborada pelo autor.

A comparação dos resultados obtidos são mostrados na Tabela 1.

Tabela 1: Comparação dos resultados

| Rede | mIoU |
|---|---|
| Tao, Sapra e Catanzaro (2020) | 0.8510 |
| Gustafsson, Danelljan e Schön (2020) | 0.4425 |
| Meu modelo (Sem *Random Crop*) | 0.5003 |
| Meu modelo (*Random Crop* 512x512) | 0.4242 |

Fonte: Tabela elaborada pelo autor.

## 4 Conclusões

Nesse projeto, foi desenvolvida uma rede neural profunda capaz de estimar as incertezas associadas às predições, no contexto da tarefa de segmentação semântica. Pela predição e incerteza da Figura 2, pode-se perceber que as regiões de maior incerteza tem uma predição mais ”granulada”, que é na região de transição entre duas classes. Mesmo que o desempenho da rede na tarefa de segmentação semântica não foi próximo aos resultados apresentados nos trabalhos do estado da arte, os valores da métrica mIoU foram um pouco abaixo dos apresentados em (GUSTAFSSON; DANELLJAN; SCHÖN, 2020). A maior vantagem do modelo apresentado é que apenas uma rede precisou ser treinada para inferir incertezas, em vez de múltiplas redes para a composição de um *ensembling*.

## Referências

CHOLLET, F. et al. **Keras**. 2015. <https://keras.io>.

CORDTS, M.; OMRAN, M.; RAMOS, S.; REHFELD, T.; ENZWEILER, M.; BENENSON, R.; FRANKE, U.; ROTH, S.; SCHIELE, B. **The Cityscapes Dataset for Semantic Urban Scene Understanding**. 2016. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

DWIVEDI, P. **Understanding and Coding a ResNet in Keras**. 2019. <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>.

GUSTAFSSON, F. K.; DANELLJAN, M.; SCHÖN, T. B. Evaluating scalable bayesian deep learningmethods for robust computer vision. **IEEE**, Junho 2020.

KENDALL, A.; GAL, Y. What uncertainties do we need in bayesian deep learning for computer vision? Dezembro 2017.

TAO, A.; SAPRA, K.; CATANZARO, B. Hierarchical multi-scale attention for semantic segmentation. 2020.

Portal de Eventos Científicos da Escola de Engenharia de São Carlos
Anais eletrônicos do 30º Simpósio Internacional de Iniciação Científica e Tecnológica da USP & 36º Congresso de Iniciação Científica e Tecnológica em Engenharia
Disponível em: http://eventos.eesc.usp.br/silicusp2022
SBi Serviço de Biblioteca Prof. Dr. Sérgio Rodrigues Fontes