



# Homework 3

## ELE2765 - Deep Learning

Thiago Matheus Bruno da Silva - 1413286

## 1 Objetivo

O objetivo do trabalho foi realizar uma segmentação semântica no dataset de de imagens de satélite do desmatamento da amazônia com a arquitetura U-net e a função de custo *weighted cross entropy*. Foi realizado três diferentes tipos de treino cada um com *patchesize* diferente e comparado os resultados. Fora isso os hiperparâmetros usados foram os mesmos. O stride utilizado foi sempre metade do número de patches, pois era a quantidade que o meu computador local aguentava alocar na memória RAM.

Model/parameters	learning rate	patche size	stride	batch size
Modelo1	1,00E-04	128	64	32
Modelo2	1,00E-04	64	32	32
Modelo3	1,00E-04	32	16	32

Figura 1: Parâmetros

## 2 Resultados

### 2.1 Gráficos

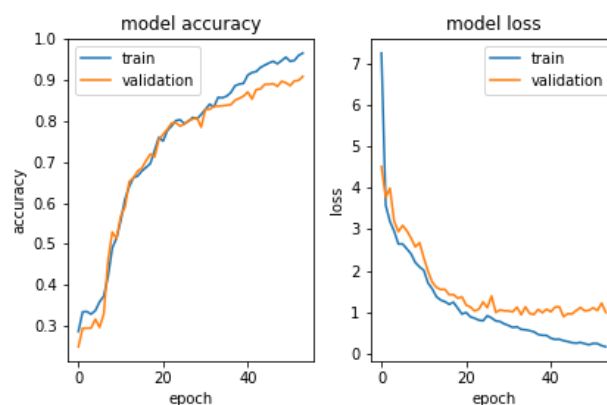


Figura 2: Modelo 1

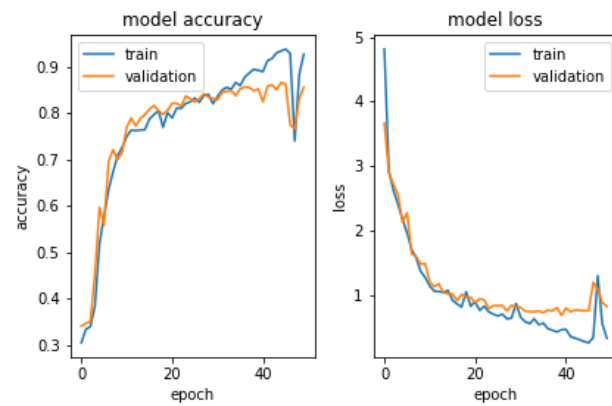


Figura 3: Modelo 2

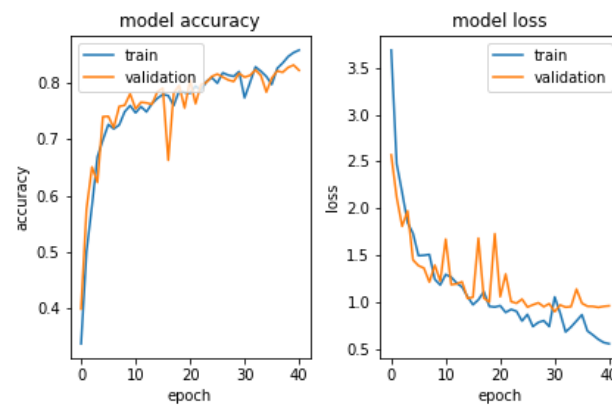


Figura 4: Modelo 3

## 2.2 Inferência

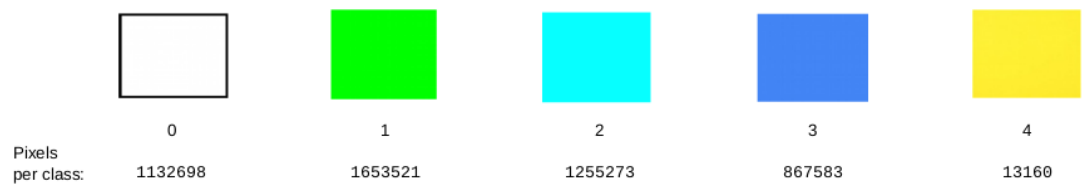


Figura 5: Cores de cada classe

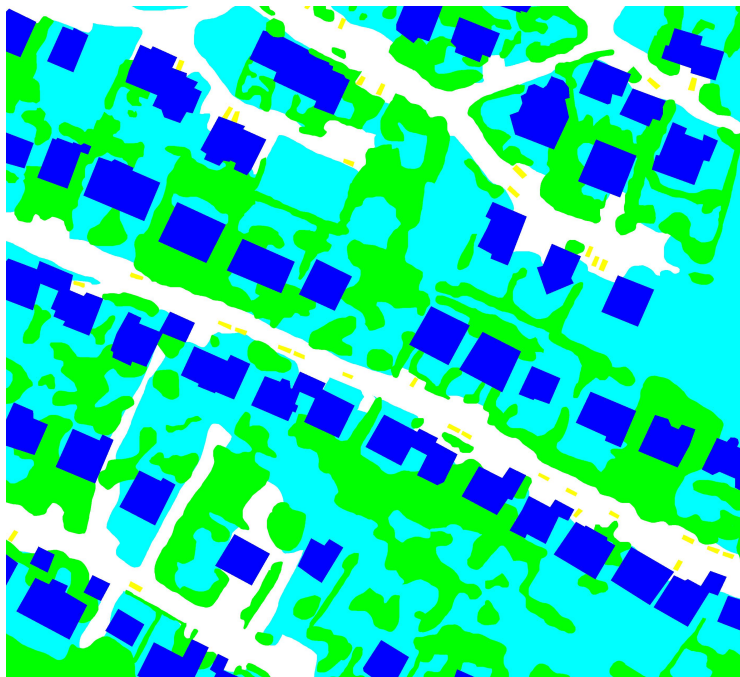


Figura 6: Referência

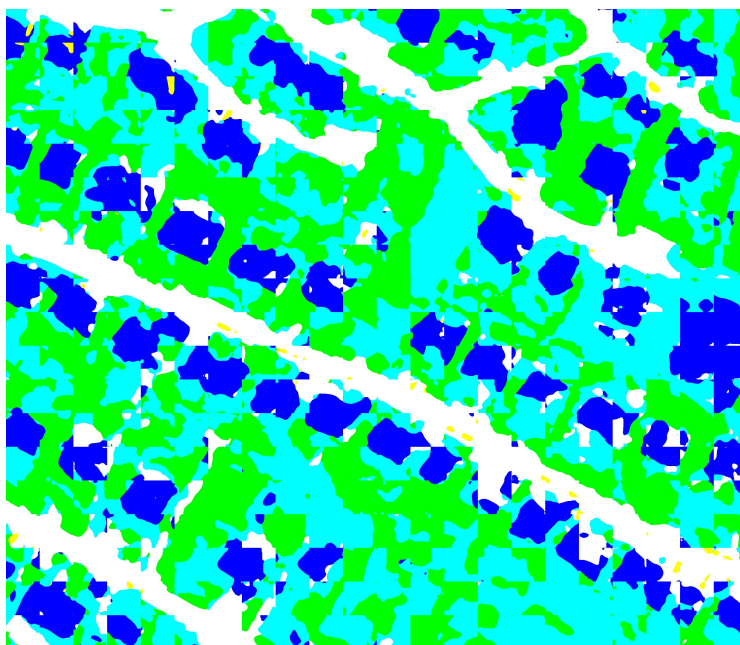


Figura 7: Modelo 1

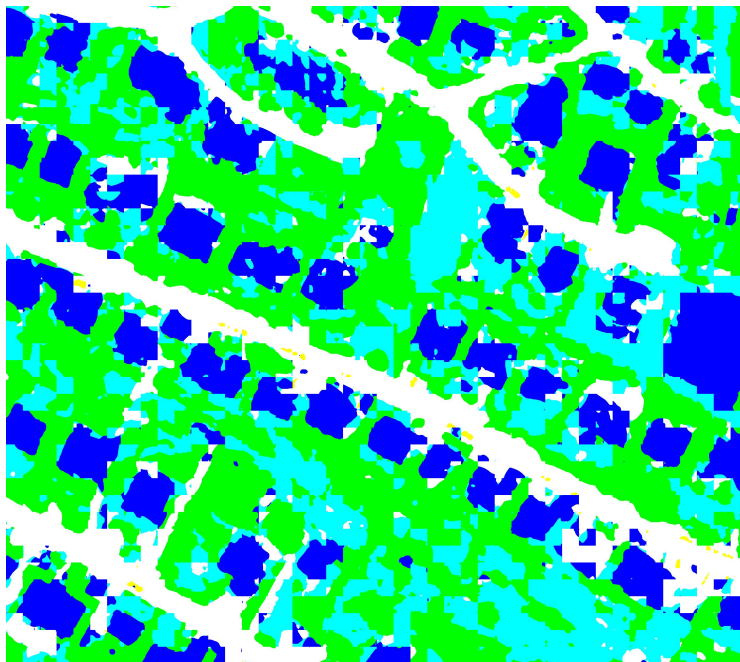


Figura 8: Modelo 2

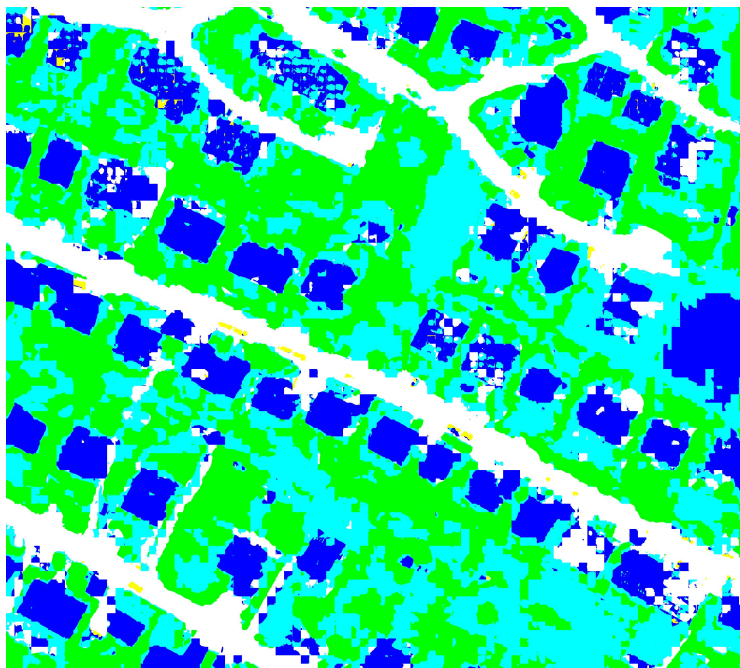


Figura 9: Modelo 3

## 2.3 Métricas

Accuracy:	74,83				
Classes:	0	1	2	3	4
F1score:	78,75	76,23	69,13	80,24	42,59
Recall:	80,77	78,79	67,10	78,86	36,57
Precision:	76,83	73,84	71,28	81,67	51,00

Figura 10: Modelo 1

Accuracy:	72,73				
Classes:	0	1	2	3	4
F1score:	76,29	77,36	61,20	79,04	40,88
Recall:	82,51	89,08	49,75	83,05	29,27
Precision:	70,95	68,36	79,49	75,40	67,92

Figura 11: Modelo 2

Accuracy:	74,08				
Classes:	0	1	2	3	4
F1score:	76,24	77,15	67,40	79,48	39,59
Recall:	77,76	81,79	62,92	80,82	33,91
Precision:	74,77	73,02	72,57	78,19	47,54

Figura 12: Modelo 3

## 3 Comentários

Como já foi dito antes, todos os modelos foram treinados usando a mesma configuração, exceto pelos diferentes tamanhos dos patches extraídos das imagens. Além disso, todos os modelos utilizaram early stopping usando a função de custo como referência. Salvando sempre o peso de menor valor na validação da função de custo e parando após 10 épocas incrementando o valor da validação da "loss".

Pelos gráficos de treinamento podemos ver que conforme diminuimos o tamanho do patch a acurácia não cresceu tanto no treino e na validação. Além disso, os gráficos com maior o maior tamanho de patches se mostram menos ruidosos que os de menores, mostrando um melhor ajuste as curvas de treino e consequentemente uma melhor generalização.

Olhando para a inferência e para as métricas fica mais nítido essa diferença de performance nos treinos. O modelo 1 apresentou a maior acurácia dentre os três modelos, e o melhor f1 score em todas as classes, que é responsável por harmonizar o recall e a precisão. Se compararmos com os outros modelos, podemos ver manchas azul claro da classe 2 em cima do azul escuro da classe 3. Também é possível identificar maiores manchas verdes correspondente a classe 1, mostrando uma piora da segmentação conforme a diminuição do patch size. Esses resultados também podem ser evidenciados olhando a acurácia referente ao modelo 2 e 3, as quais são menores que o modelo 1. Podemos ver também, que a precisão referente as casses 1 e 3 caíram nos modelos 2 e 3, ou seja, a taxa de falso positivos

aumentou, confirmando o observado na imagem de inferência.

Outra questão importante se refere a quantidade de pixels em cada classe no dataset. Como utilizamos a função de custo **weighted categorical cross entropy** para balancear as classes com poucos pixels, no caso a 3 e 4. A terceira, possui um pouco menos pixels classificados do que as classes restantes, enquanto a 4 possui duas ordens de grandeza a menos. Fica evidente que o balanceamento feito pela função de custo implementada ajudou a classificação da classe 3, mostrando resultados melhores que o da classe 2, a qual possuía mais amostras em todos os modelos. Entretanto, a classe 4, cor amarelo, não conseguiu bons resultados devido a pouca quantidade de pixels no nosso dataset. Mesmo com a função de custo ponderando as classes com base na quantidade de pixels de cada uma não foi o suficiente para dar resultados convincentes a última classe.

## 4 Conclusão

Como foi dito antes, o modelo 1 obteve os melhores resultados, os quais foram piorando gradativamente conforme diminuimos o patch size. Podemos explicar tal fenômeno pela perda de contexto global de cada imagem. A primeiro momento podemos pensar que a diminuição do tamanho de cada patch irá consequentemente mais imagens e por isso melhorará a acurácia da rede. Porém, essa diminuição de tamanho acarreta numa perda de contexto da rede, classificando erradamente os pixels. Para classificarmos a imagem toda, cuja resolução é muito alta precisamos dividir em vários pedaços a imagem, porém esses pedaços não podem ser tão pequenos de forma que comprometa o contexto global da imagem.