Image Segmentation

Anders Henriksen	s183904
Asger Schultz	s183912
Oskar Wiese	s183917
Mads Andersen	s173934
Søren Winkel Holm	s183911

December 21, 2019

Contents

1 Abstract

2 Introduction

2.1 Brief overview of data

Our data consist of one large, high resolution orthomosaic photo of a sugar cane field formatted as a RGB image. The field has been manually labelled by expert biologist to create a human-ground truth, these labels are represented as a GT-matrix with 3 possible classes. Each pixel is either classified as a crop row (green), weed (yellow), background (red). The large photo is cropped into smaller images and afterwards augmentation techniques are used to gain more data. We will return to the augmentation part later. The images with exclusively black pixels are removed from our data set, and images with some black pixels are ignored when calculating the loss function

3 Methods

Netværket blev initialiseret

The encoder part of the network creates a rich feature map representing the image content. The more layers of max-pooling there are the more translation invariance for robust classification can be achieved. The boundary detail is very important when dealing with image segmentation. Hence, capturing boundary information in the feature maps of the encoder before upsampling is important. This can simply be done by storing the whole feature map, but due to memory constrains only the maxpooling indices are saved, which is a good approximation of the feature maps.

3.1 Loss function: Quality over quantity

Multi-class cross entropy because:

- Softmax Network: Minus log likelihood
- Can be seen as a classic multiclass classifier – just on a pixel-by-pixel basis.

Weighted cross entropy because:

- Unbalanced class distribution: Network has to learn to focus on important pixels: Don't classify everything as dirt.

- Initial tests made the network behave as the baseline: Simple features in early layers got were not penalized enough and learning was not stable.
- Resampling expensive

3.2 Metrics

Had to use different metrics because

- Not agreement in Image Segmentation papers.
- Want to get accuracy on a global scale and on a class scale.
- Different metrics important in different fields.

The metrics [1][2]

- Global accuracy: Trivial and not very important because of class imbalance but is good for smoothness
- Mean class-wise accuracy: Takes class imbalance into account. Is what is being optimized for in the model.
- Mean Intersect over Union: "Jaccard Index". Found to be better correlated with human classification though still only ≈ 0.5. Favours region smoothness highly and not boundary accuracy.
- Harmonic mean of precision and recall. To compare to others with same project. Penalizes false positives and gives less credit to true negatives thus being better for unbalanced classes.

3.3 Regularization and Hyperparameters

Regularization

- NN's are prone to overfitting, because they are so flexible
- Prevent overfitting → better results on test data
- Three methods

[1] https://hal.inria.fr/hal-01581525/document

[2] http://www.bmva.org/bmvc/2013/Papers/paper0032/paper0032.pdf

- Dropout: Randomly remove nodes to increase variability. $p = 10\,\%$
- Data augmentation: Increase size of dataset
  - Crop each 512×512 to random 256×256
  - $50\,\%$chance of flip T/D and $50\,\%$chance of flip L/R
- Batch normalization normalizes activations
  - Faster convergence
  - Prevents ReLU from not learning
  - Introduces noise
  - Reduces vanishing/exploding gradient problem, as values stay close to 0

Hyperparameters

- Adaptive learning rate from ADAM optimizer, initialized at $2 \cdot 10^{-4}$
- Total: 26 conv + batchnorm + ReLU with dropout, 5 pool/upsample, 1 softmax
- 14.7 M parameters in encoder – significantly lower than 134 M in VGG16 because of no fully-connected layers
- Kernel size: 3×3, stride 1, maxpool: 2×2, stride 2
- Corresponding padding of 1 to prevent reduction of image size

https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-tech
https://medium.com/deeper-learning/glossary-of-deep-learning-batch-normalisation-8266dcd2fa

3.4 Unification of cropped image predictions

In a real-world application of the segmentation a farmer would want a complete and precise segmentation of his whole field at once, such that fertilization and pesticides can be distributed accordingly. However, it turns out that the prediction quality is of less quality near the boarders of the image. Therefore, a naïve stitching of the cropped images leads to a full-blown image prediction with obvious flaws near the boarders between the cropped images. To solve this problem, we have chosen to increase the size of the cropped images, and infer on these enlarged pictured. In the procedure of joining the enlarged cropped pictures the pictures are cropped again, to avoid the near border areas. This is computationally inefficient, but it works, and since the inference time is not that big, it is an alright solution. For industrial purposes, another approach might be beneficial.

4 Results

5 Discussion

5.1 Comparison of different image segmentation neural networks

- Several competing network structures with high performance in image segmentation. U-net, FCN, DeepLabv1, DeconvNet
- Purpose of SegNet, efficient
- 3 out of the 4 mentioned uses the encoder from the famous VGG16 paper, but differ in decoder.
- FCN, No decoder -¿ Blocky segmentation, but very efficient in inference time.
- DeconvNet, Deconvolution and fully connected layers.
- U-Net, (different purpose), skip connections.
- Main takeaway
- (Deeplabv-LargeFOV & FCN)

5.2 Extension of network

6 References

[1] Kendall, Alex et al.: ”SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”. PAMI, 2017.

[2] Wangenheim von, Aldo et al.: ”Weed Mapping on Aerial Images”. INCoD.LAPIX.01.2019.E.

7 Appendix