

### *Image segmentation with saliency maps*

Tudor Buzu

## Problem definition

In this project, I combine different saliency map generation techniques with a neural network (UNet) to see if the final results of image segmentation can be improved.

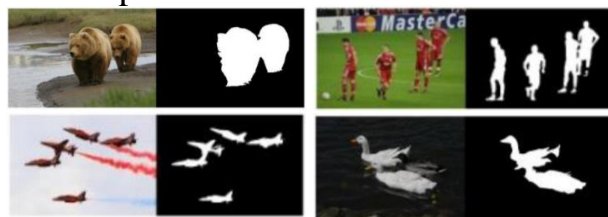
### Work plan:

- Select a pretrained CNN (VGG16);
- Choose 2 image segmentation datasets (CMU-Cornell iCoseg, Inria Aerial Image Labeling);
- Compute saliency maps (three approaches: backprop, deconv, guided backprop) on the selected datasets;
- Combine saliency maps with a UNet;
- Evaluate final results using IoU, Dice.

## Datasets

### 1.) CMU-Cornell iCoseg Dataset (subset\_80)

- a subset of 80 images containing people, airplanes and animals.



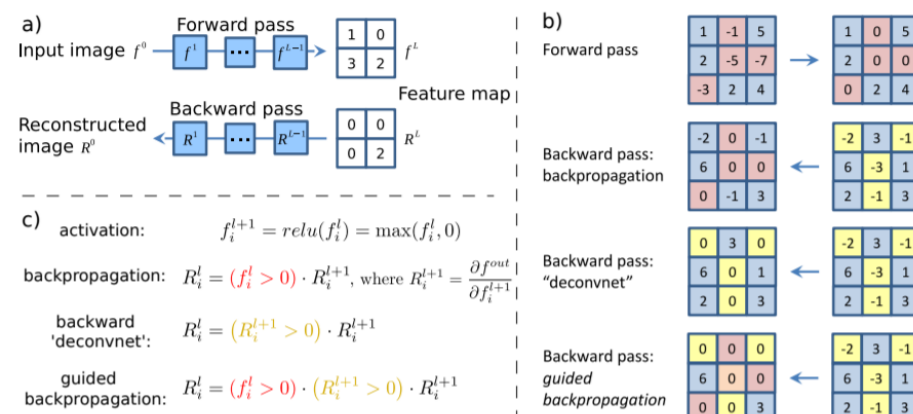
## 2.) Inria Aerial Image Labeling Dataset (subset\_chicago)

- A subset of 400 aerial images with ground truth data for two semantic classes: *building* and *not building*.

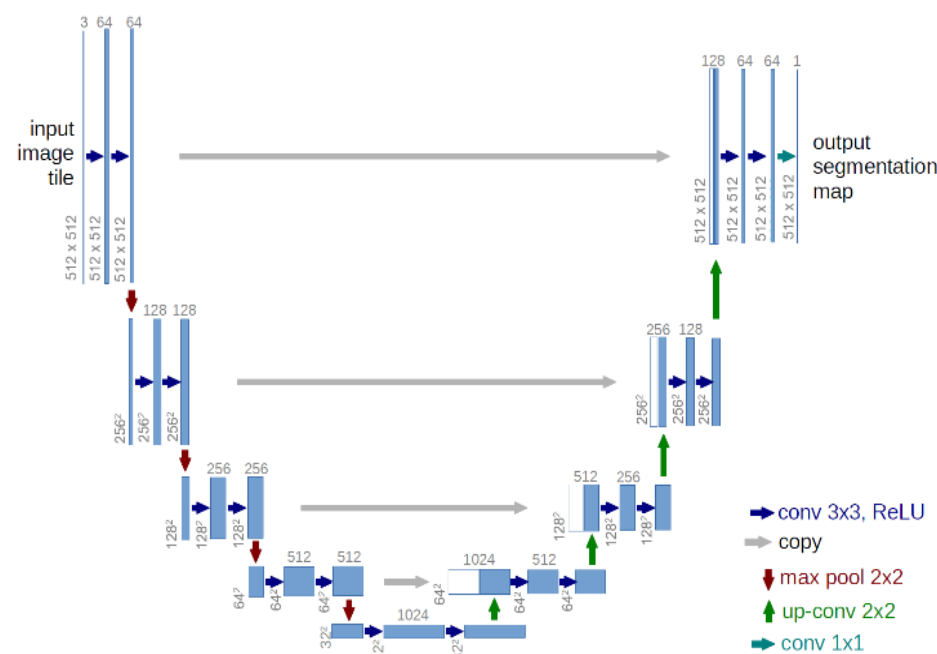


## Saliency map approaches

- 1.) Backprop modifier = None: propagate gradients only for positive activations;
- 2.) Backprop modifier = Deconv: propagate only positive gradients;
- 3.) Backprop modifier = Guided: propagate only positive gradients for positive activations.

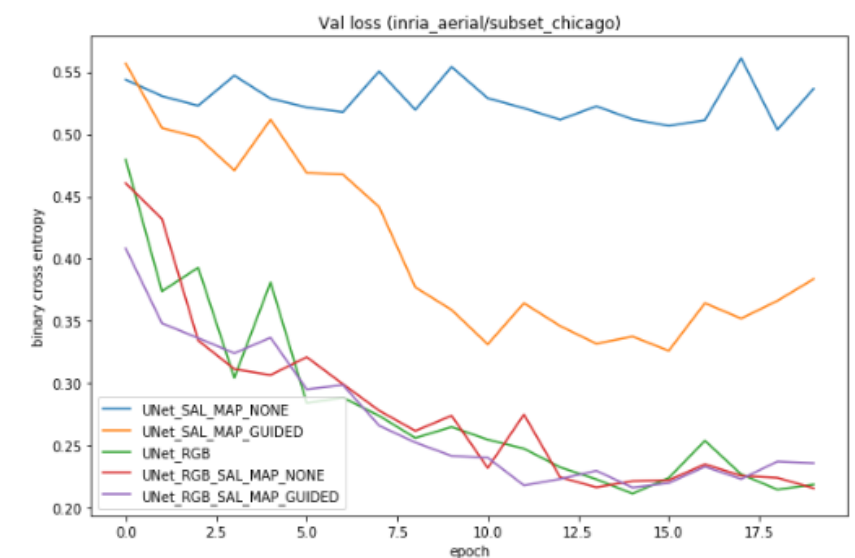
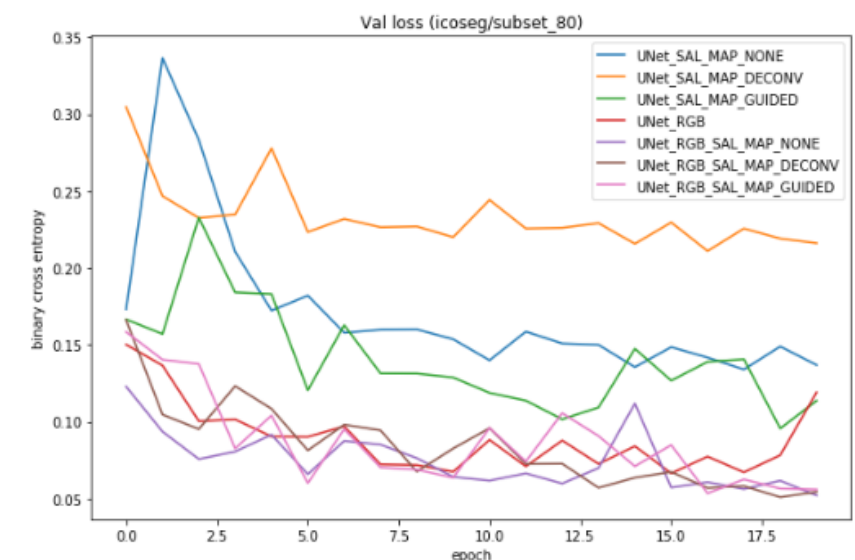


## UNet architecture

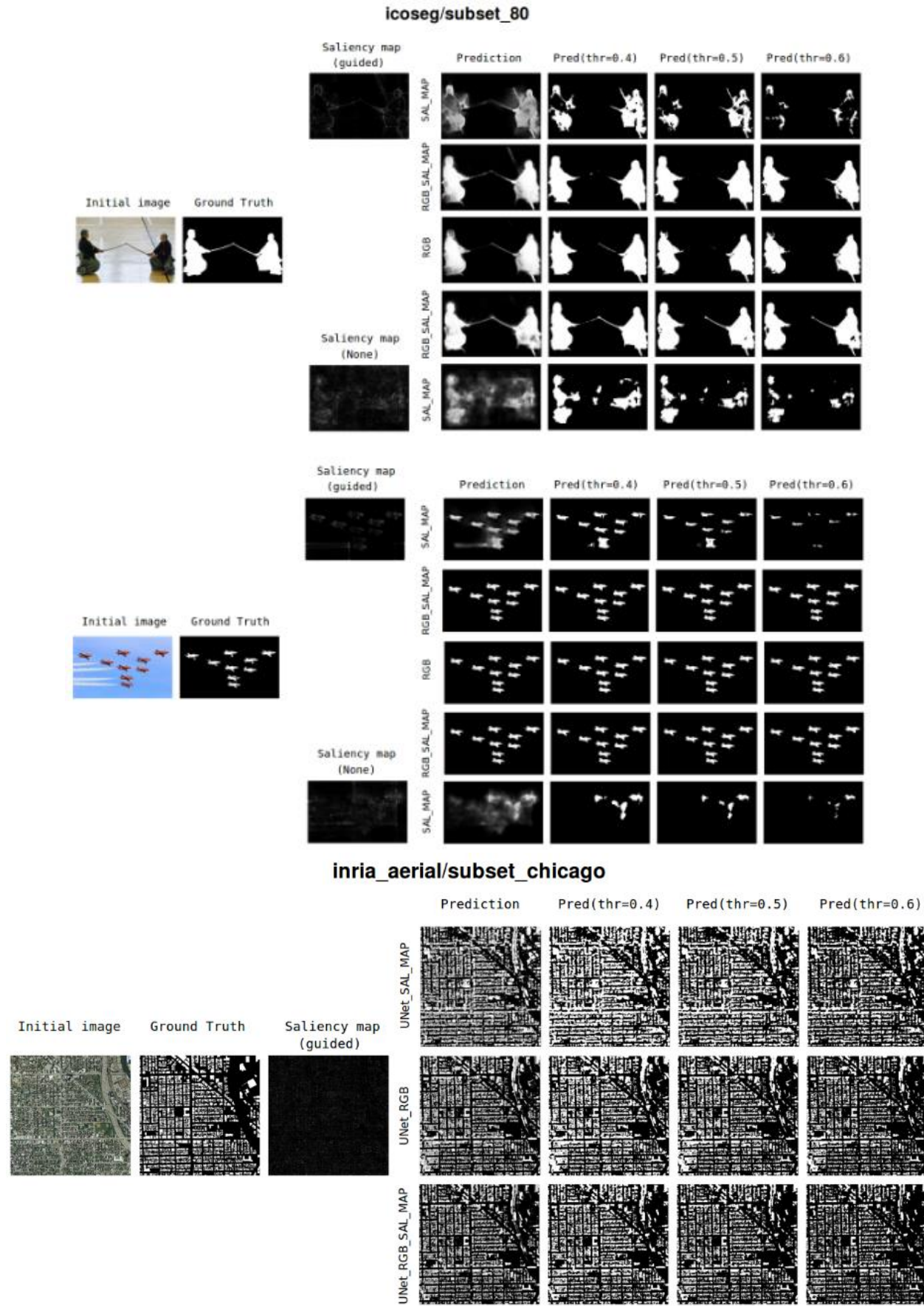


## Training details

- **Loss:** binary cross entropy;
- **Optimizer:** Adam(lr=1e-4);
- **Metrics:** IoU and Dice;
- **Prediction\_threshold:** 0.5;
- **Num\_epochs:** 20;
- **Batch\_size:** 1;
- **UNet\_input\_shape:** (512, 512, ch), ch={1,3,4}.



# Qualitative results



# Quantitative results

Dataset name	Experiment name	Train		Val		Test	
		IoU	Dice	IoU	Dice	IoU	Dice
icoseg/subset_80	UNet SAL MAP NONE	0.178	0.273	0.174	0.281	0.184	0.283
	UNet SAL MAP DECONV	0	0	0	0	0	0
	UNet SAL MAP GUIDED	0.405	0.561	0.412	0.563	0.407	0.562
	UNet RGB	0.723	0.814	0.642	0.720	0.653	0.736
	UNet RGB SAL MAP NONE	0.733	0.822	0.724	0.796	<b>0.719</b>	<b>0.796</b>
	UNet RGB SAL MAP DECONV	0.743	0.830	0.687	0.771	0.693	0.782
	UNet RGB SAL MAP GUIDED	0.702	0.800	0.681	0.767	0.697	0.780
inria_aerial/subset_chicago	UNet SAL MAP NONE	0	0	0	0	0	0
	UNet SAL MAP GUIDED	0.630	0.773	0.532	0.695	0.462	0.632
	UNet RGB	0.734	0.847	0.674	0.805	0.549	0.709
	UNet RGB SAL MAP NONE	0.754	0.860	0.672	0.804	0.552	0.712
	UNet RGB SAL MAP GUIDED	0.744	0.853	0.675	0.806	<b>0.568</b>	<b>0.725</b>

# Conclusions

- Saliency maps contain enough information to improve semantic segmentation of both natural and aerial images.
- On the iCoseg subset, saliency maps with standard backpropagation (None) improve the results with ~6% (IoU and Dice). On Inria aerial images, saliency maps with guided backpropagation improve the results with just ~1.5% (IoU and Dice).
- It is interesting to mention that saliency maps on the iCoseg subset make the network more confident at the margins of the objects, while on the Inria aerial images help to reject roads and everything that is from the “*not building*” class.

# References

- “Visualizing and Understanding Convolutional Networks” (Zeiler and Fergus, 2013)
- “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps” (Simonyan et al., 2014)
- “Striving for Simplicity: The All Convolutional Net” (Springenberg et al., 2015)
- “U-Net: Convolutional Networks for Biomedical Image Segmentation” (Ronneberger et al., 2015)
- “Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark” (Maggiori et al., 2017)