# Deep learning for football video analysis

- Michał Warchalski, Data Scientist at Nordeus

# About me

- Research at ML&AI Team at Nordeus
- PhD in Mathematics at the University of Bonn

# About Nordeus

**FOUNDED**
2010

**HQ**
Belgrade, Serbia

**GAMES**
Top Eleven
Golden Boot

**CREW**
170 People,
21 Nationalities

# About our games

- Top Eleven: over 190 million registered users


- Golden Boot: 30 million played so far

# Presentation plan

- Machine learning at Nordeus.
- Technology we use.
- Further results and future challenges.

# Machine learning at Nordeus

# Machine learning for games

### FORECASTS

*By game/cohort:* Trends

### DESCRIPTIVE MODELS

*By user:* Clustering, segmentation, playing styles, ...

### PREDICTIVE MODELS

*By user:* Churn prediction

### CRM

*By user:* A/B tests

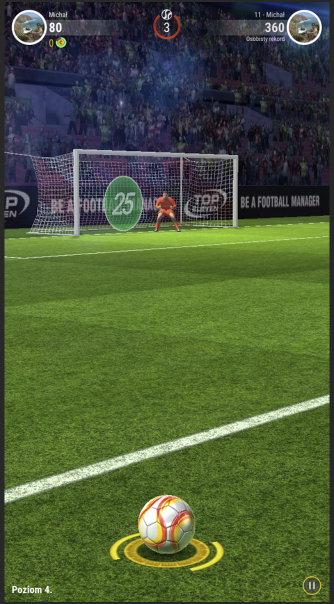
### RECOMMENDATION ENGINES

*By user:* recommend players, items, features, teammates, matchmaking, ...

# What our team does

- Apply deep learning to improve gameplay in our games.
- AI for games
- Recently: instance segmentation for mapping frames from football videos to a 2D model of football pitch

# Case study: Golden Boot

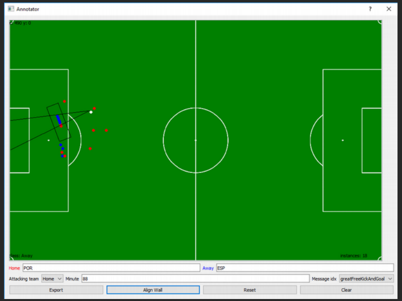





Goal of the game: score from a free kick, possibly rotating the ball to avoid the wall and the goalkeeper.

# Case study: Golden Boot - our goal

Goal: Free kick in live match delivered to millions of players in real-time.



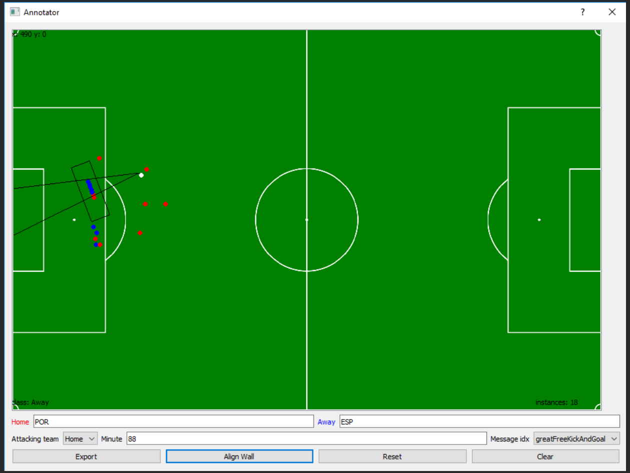
1. Live match: free kick

2. Our pipeline: 2D positions

3. Free kick in Golden Boot

# Pipeline: player detection + clustering + positions


Free kick in a football game live on TV




Positions of players for Golden Boot

# Goal: Instance segmentation

Source: http://cs231n.stanford.edu/

Source: http://cs231n.stanford.edu/

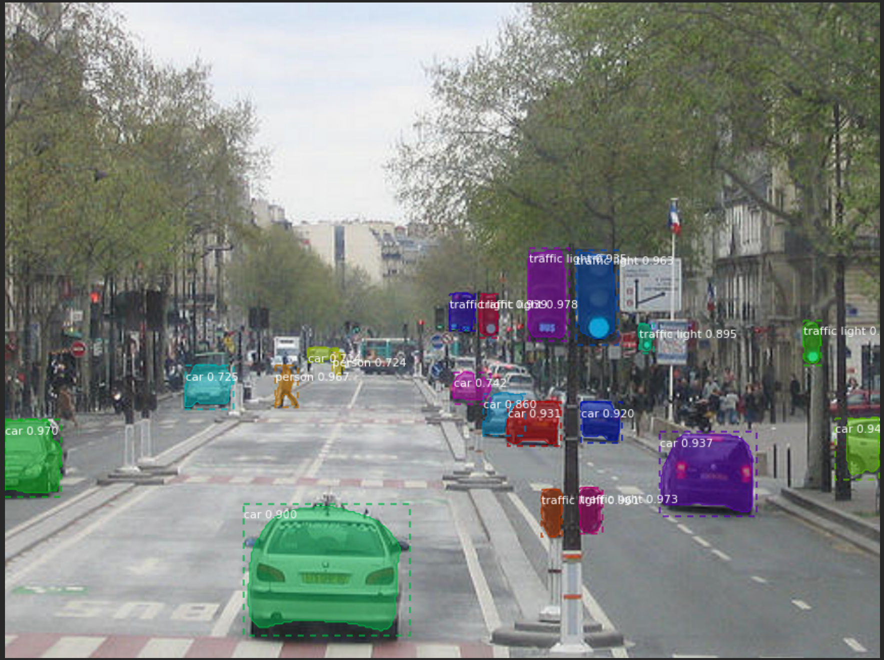
# Technology we use

# Solution: Mask R-CNN

Mask Region Convolutional Neural Network

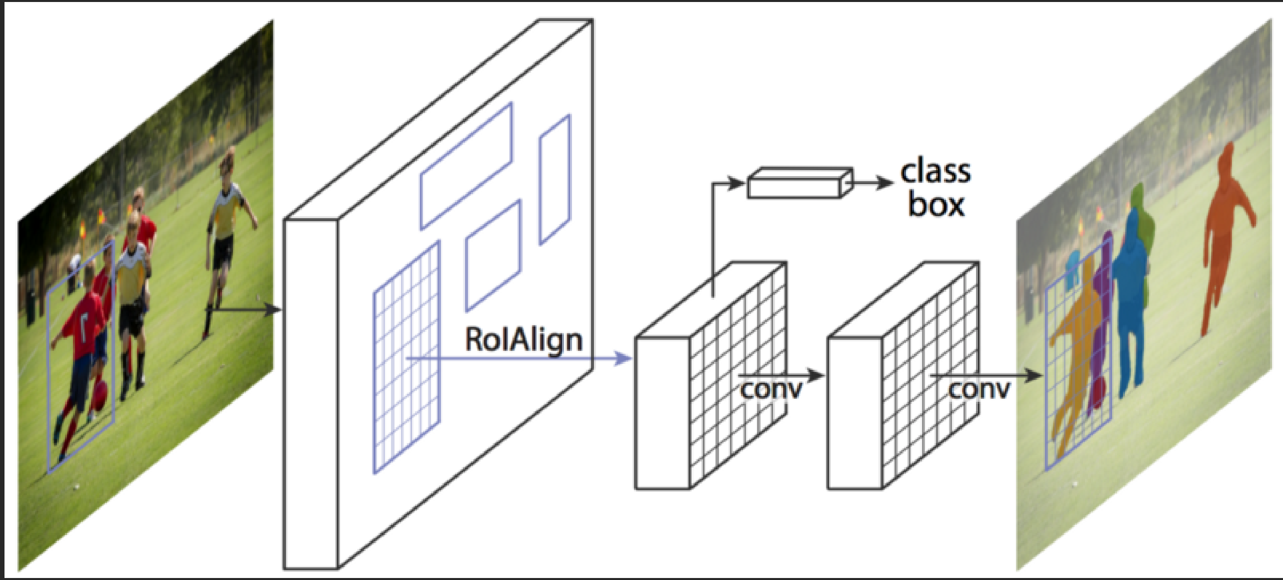
[K. He, G. Gkioxari, P. Dollár, R. Girshick ‘17]

is a framework for object instance segmentation.




Source: https://github.com/matterport/Mask_RCNN

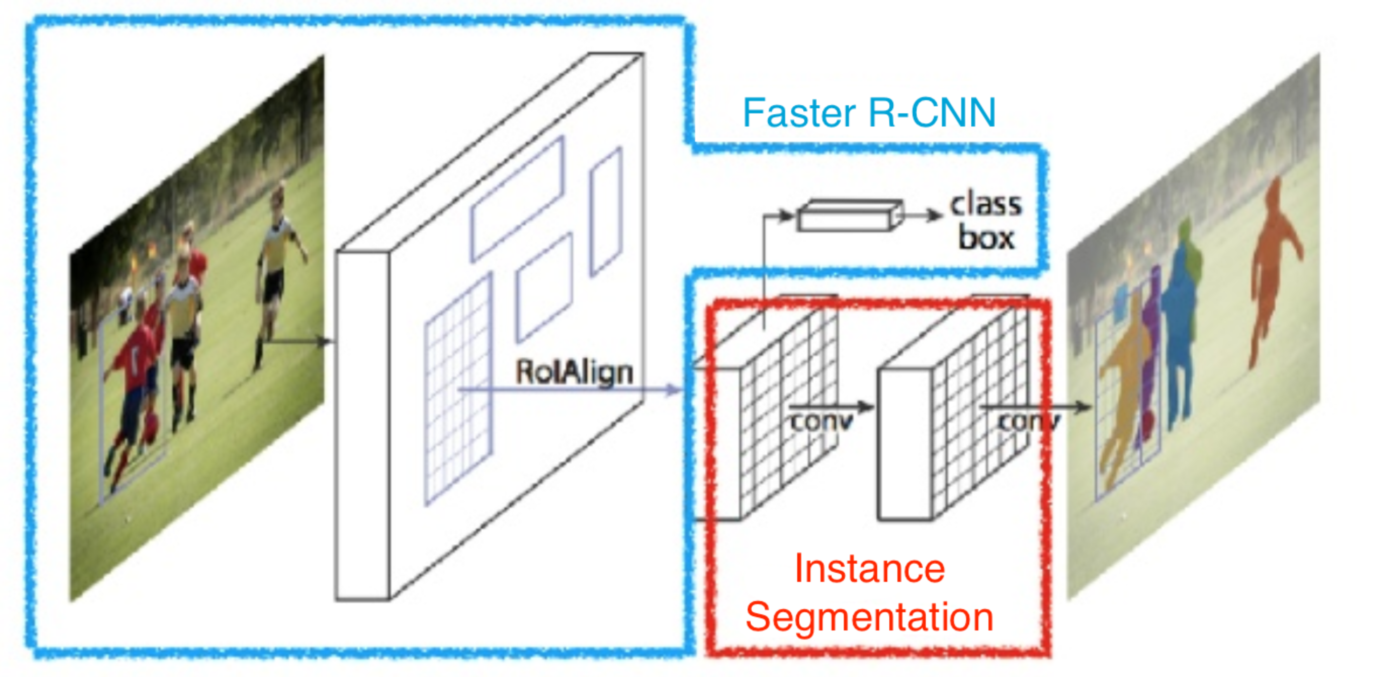
# Architecture summary




Source: *Mask R-CNN*, Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick

Feature extraction + Region Proposal + Region classification + Mask prediction

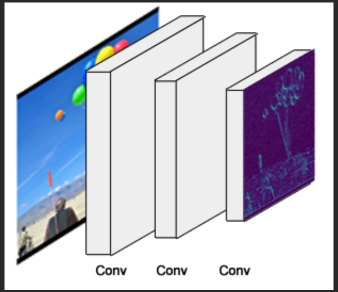
# Architecture summary



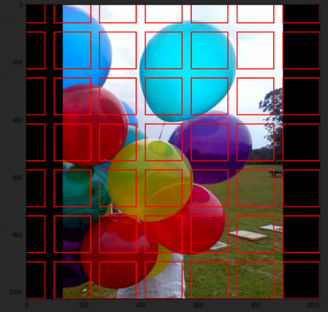

Feature extraction + Region Proposal + Region classification + Mask prediction

# Architecture: ResNet + RPN

- Feature extraction: use a CNN (ResNet) to extract features.

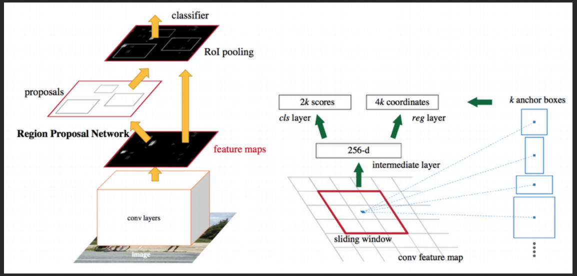



Source: https://engineering.matterport.com

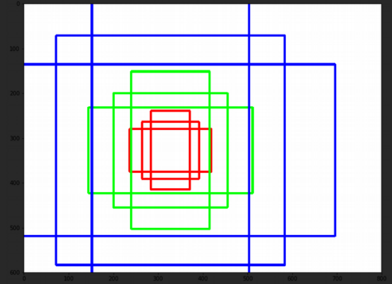


Source: https://engineering.matterport.com

- RPN: Faster R-CNN [S. Ren, K. He, R. Girshick, J. Sun ‘15]: proposal regions with a neural net.

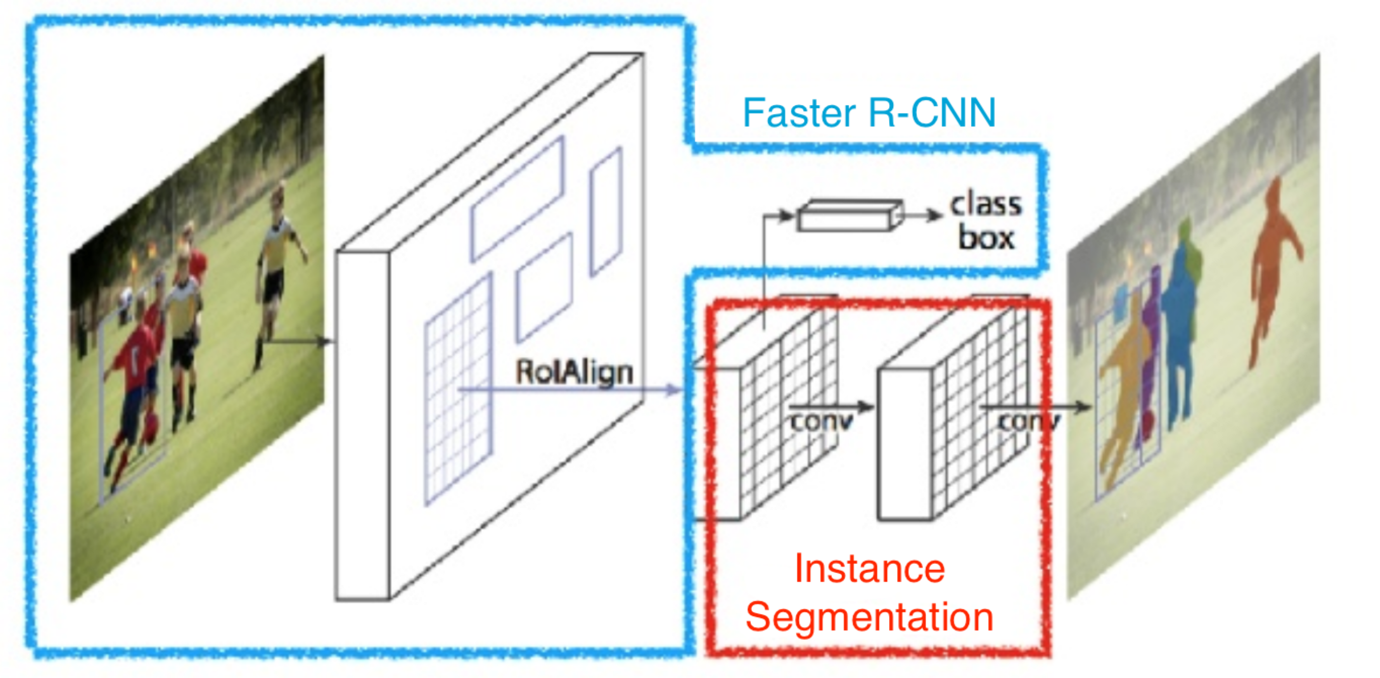



Source: Faster R-CNN: *Towards Real-Time Object Detection with Region Proposal Networks*, Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun



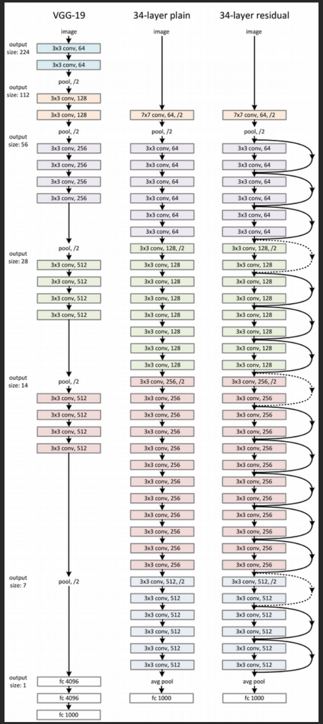
Source: https://medium.com/@smallfishbigsea

# Feature extraction

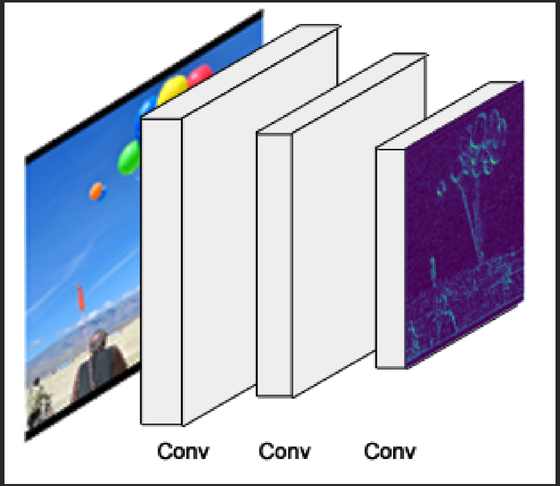


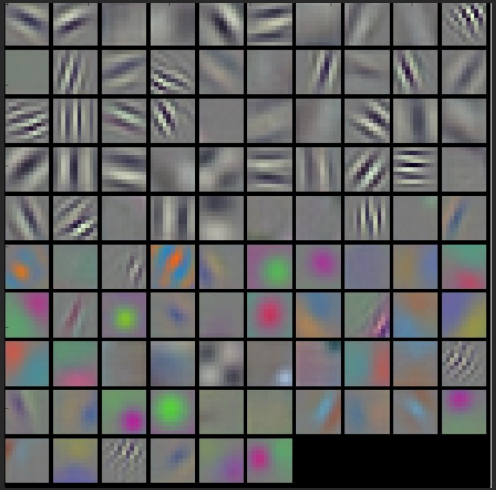

Feature extraction + Region Proposal + Region classification + Mask prediction

# Feature extraction: pre-trained CNN




Source: *Deep Residual Learning for Image Recognition*, Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

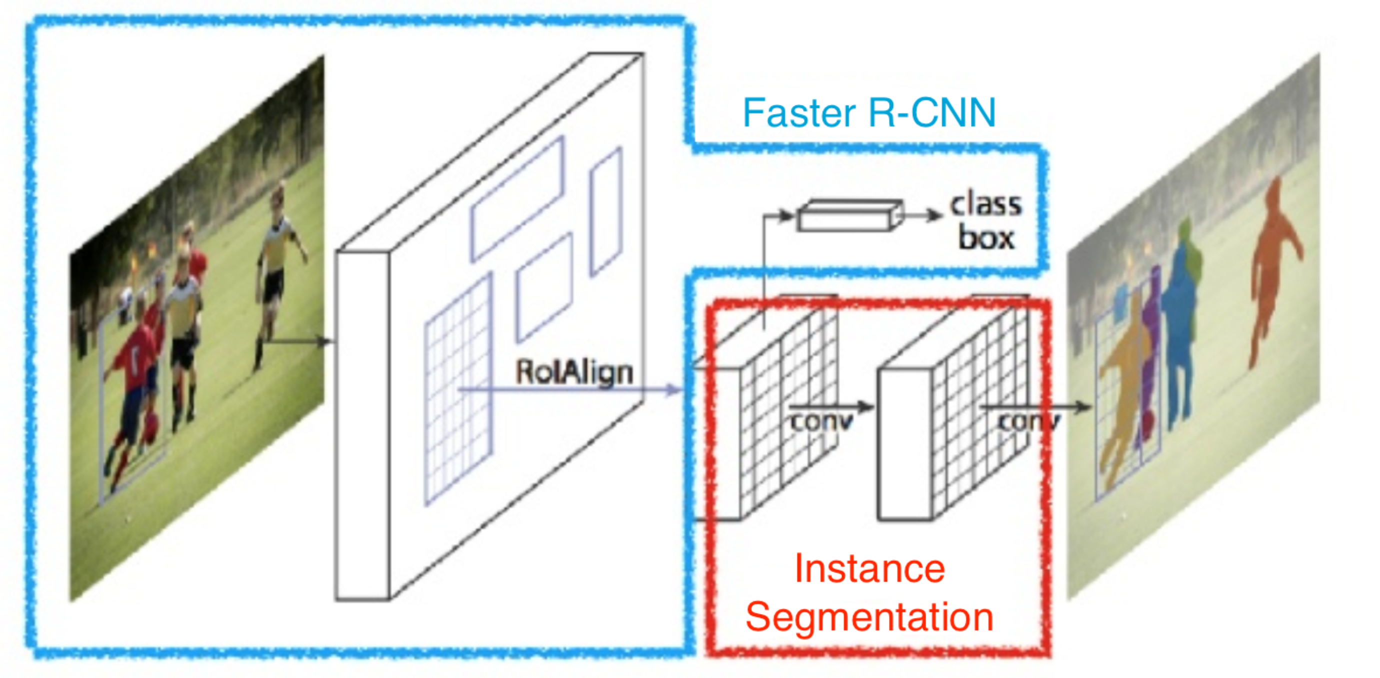


Source: First conv layer of trained AlexNet, http://cs231n.github.io/

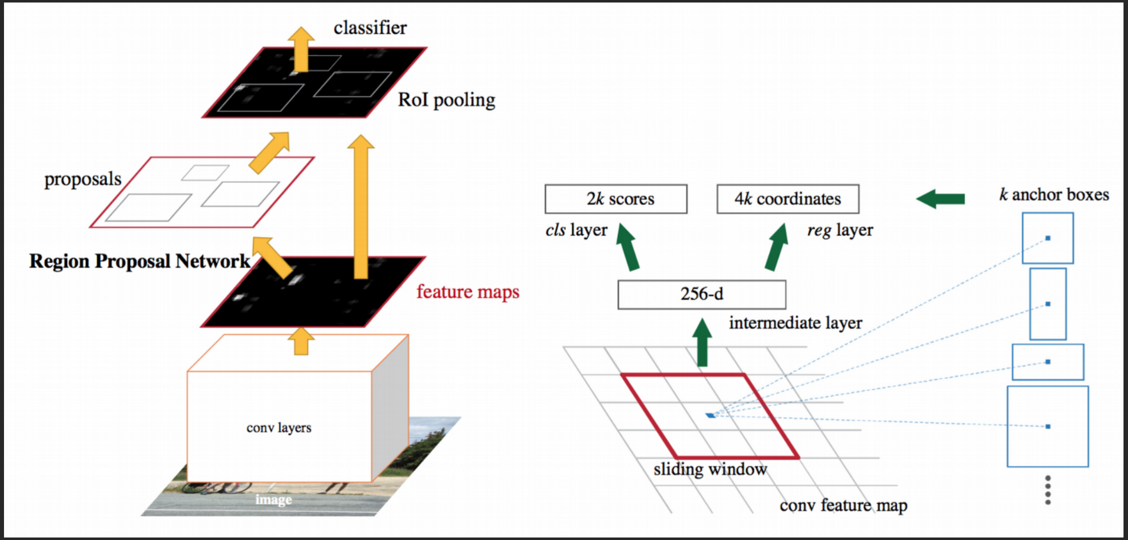
Use initial layers of a CNN pre-trained for image classification in order to extract features useful in further stages.

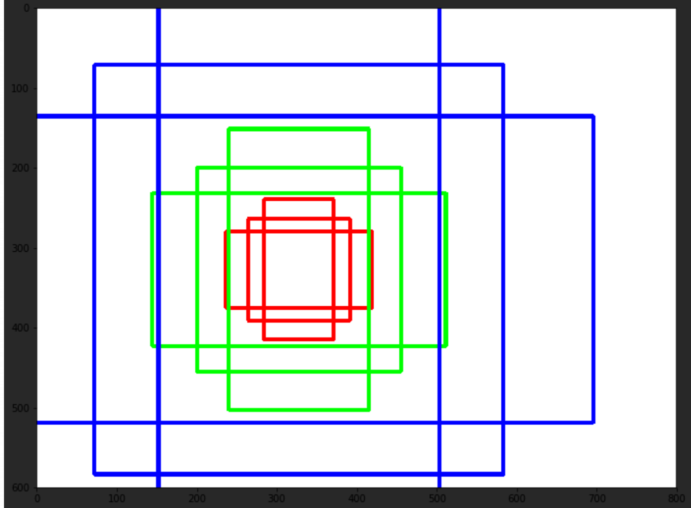
# Region Proposal




Feature extraction + Region Proposal + Region classification + Mask prediction

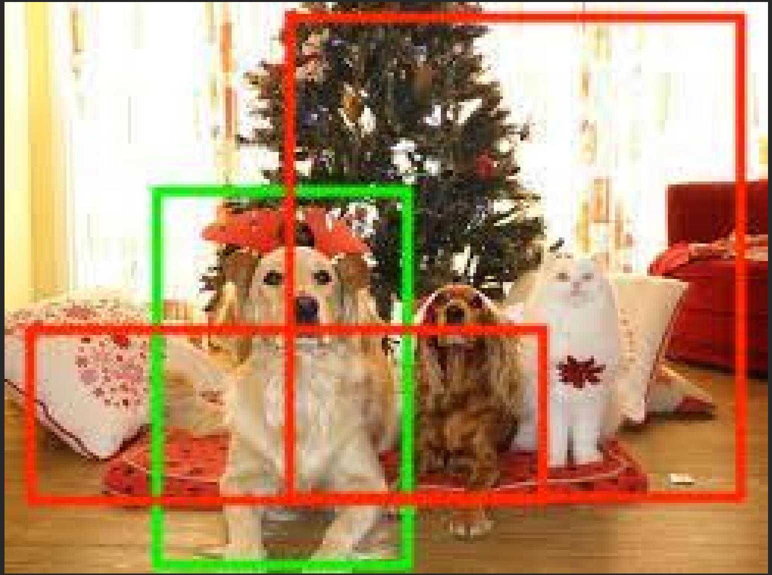
# Region Proposal Network




Slide a small window over the feature map and predict rectangles of various scales and aspect ratios, classify background vs. foreground and regress (positive boxes according to IoU).

# Good vs. bad bounding boxes



# Region Proposal Network: Loss

- Total loss:

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{box}$$

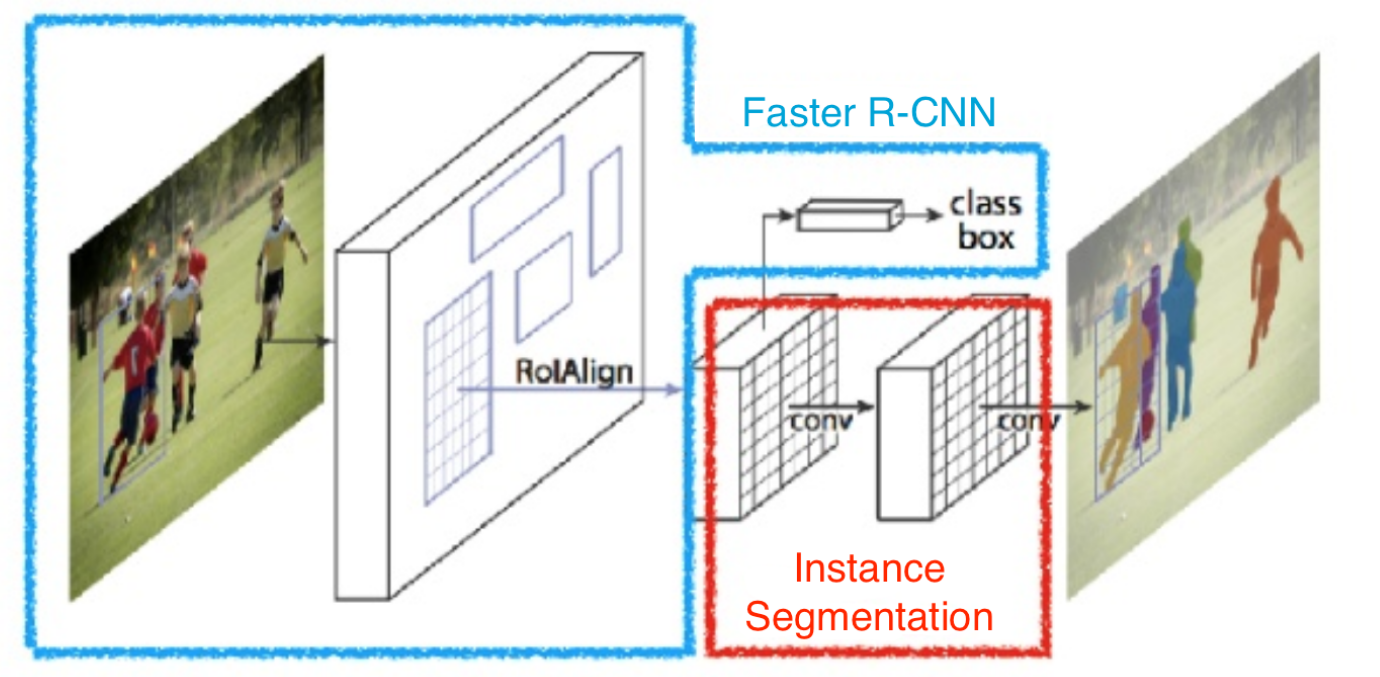
- Classification loss:

$$\mathcal{L}_{cls} = \frac{1}{N_{cls}} \sum_i -p_i^* \log p_i - (1 - p_i^*) \log(1 - p_i)$$

- Regression loss:

$$\mathcal{L}_{box} = \frac{\lambda}{N_{box}} \sum_i p_i^* L_1^{smooth}(t_i - t_i^*)$$

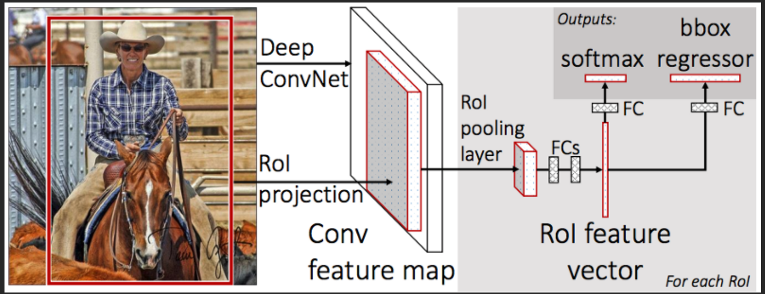
# Architecture summary




Feature extraction + Region Proposal + Region classification + Mask prediction

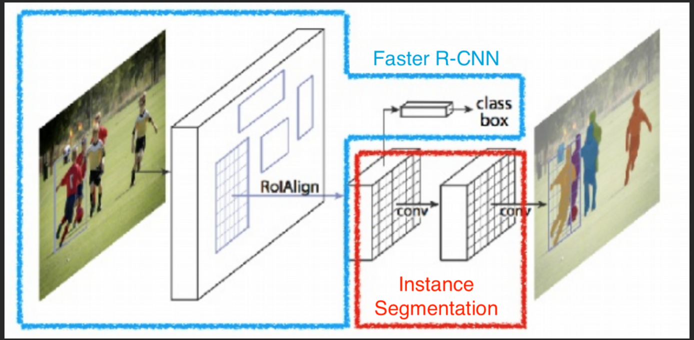
# Architecture: RoI classification + Mask prediction

- Region of Interest classification. Apply a CNN to every proposed region and classify it.



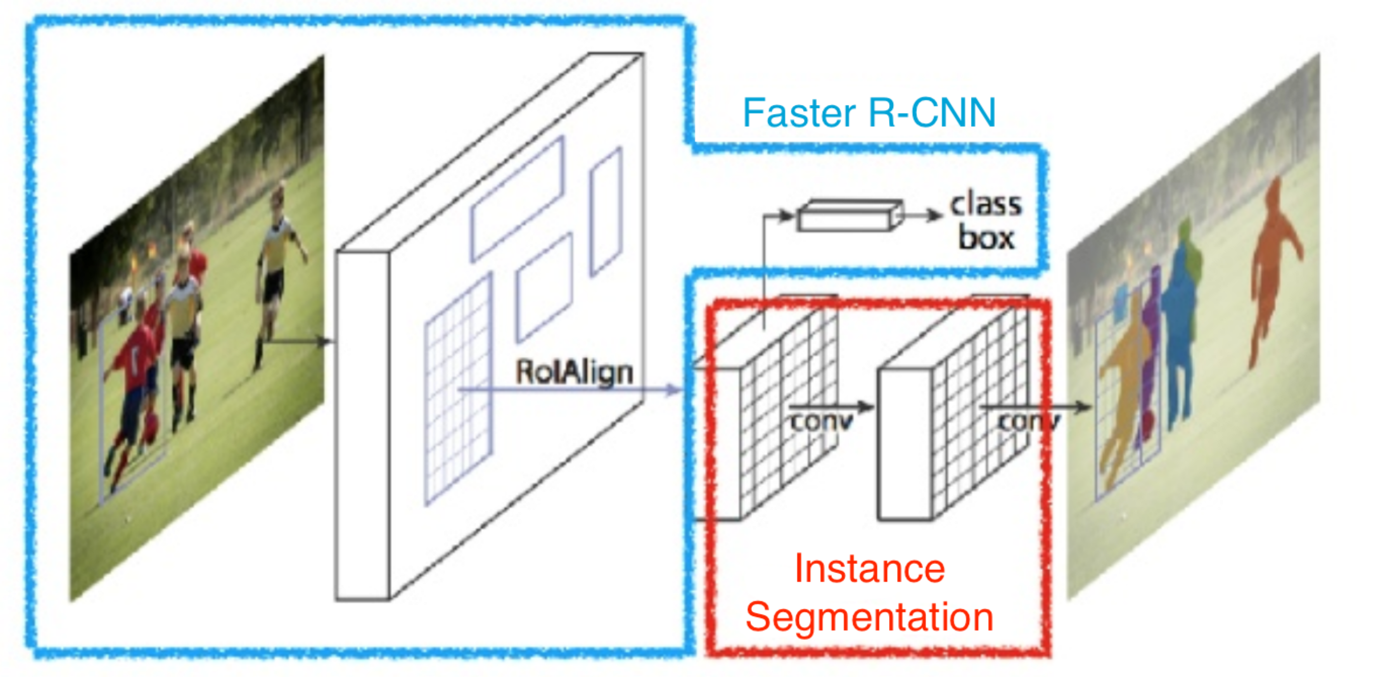

Source: Fast R-CNN, https://arxiv.org/abs/1504.08083

- Mask prediction. Classify each pixel of the proposed region classified as K, as either K or not K.



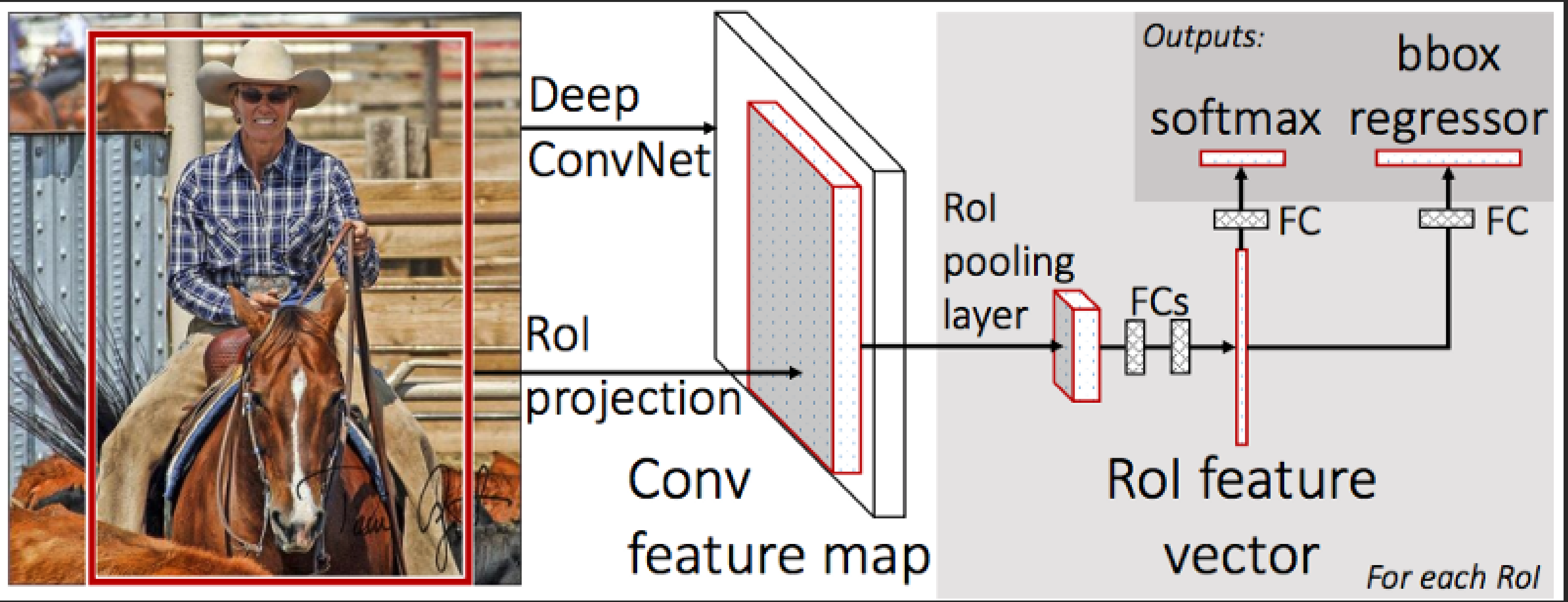

Source: *Mask R-CNN*

# Region of Interest classification




Feature extraction + Region Proposal + Region classification + Mask prediction

# Region of Interest classification




Feed the proposed region into a NN, classify it and correct the bounding box.

# Region of Interest classification: Loss

- Total loss:

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{box}$$

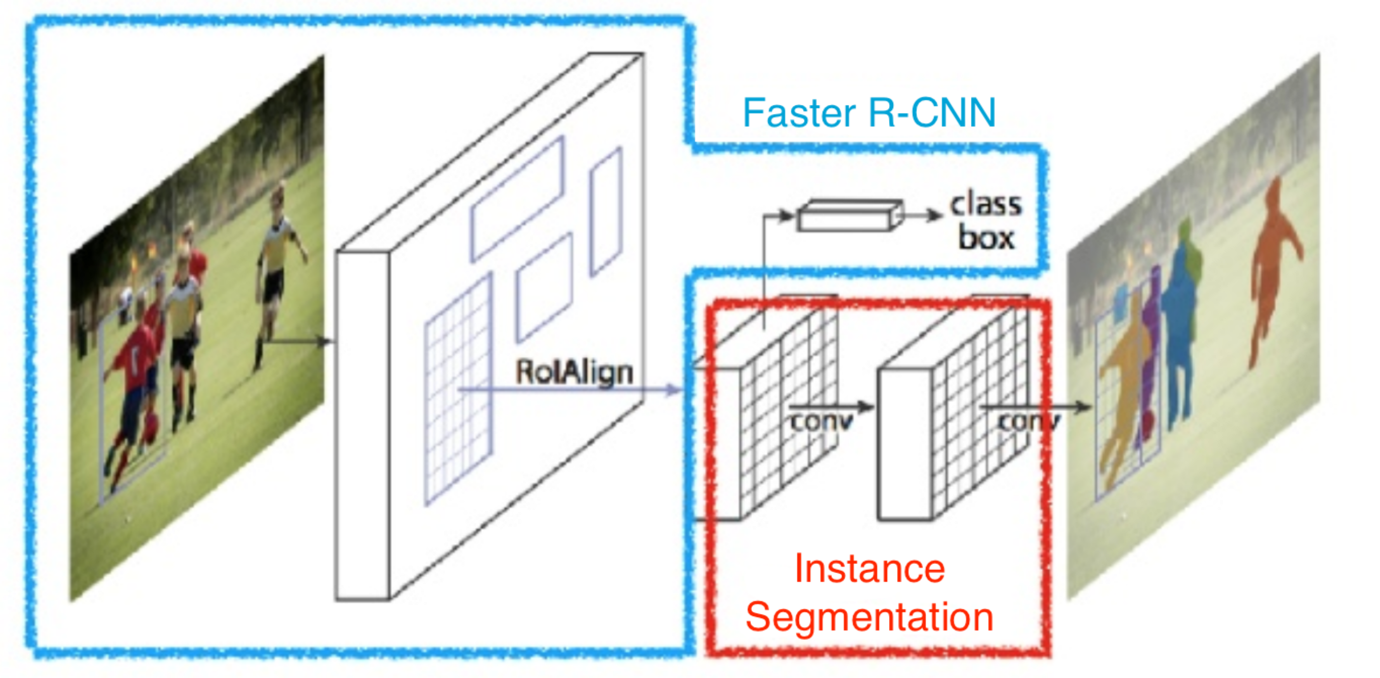
- Classification (softmax) loss, where $u_i$ is the true class label:

$$\mathcal{L}_{cls} = \frac{1}{N_{cls}} \sum_i -\log p_{u_i}$$

- Regression loss:

$$\mathcal{L}_{box} = \frac{1}{N_{box}} \sum_i \chi_{u_i \geq 1} L_1^{smooth}(t_i - t_i^*)$$

# Mask prediction




Feature extraction + Region Proposal + Region classification + Mask prediction

# Mask prediction

Given region classified as K, per pixel classification object vs. non-object

# Mask prediction: Loss

- Loss:

$$\mathcal{L} = \mathcal{L}_{mask}$$

- Mask loss (average over such) for a mask of size m x m for some class K:

$$\mathcal{L}_{mask} = -\frac{1}{m^2} \sum_{1 \leq i,j \leq m} y^*_{ij} \log y_{ij} + (1 - y^*_{ij}) \log(1 - y_{ij})$$

# Object detection: timeline

Very quick progress over last few years!

- *Rich feature hierarchies for accurate object detection and semantic segmentation*, Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik ‘13
- *Fast R-CNN*, Ross Girshick ‘15
- *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, Ross Girshick ‘15
- *Mask R-CNN*, Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun ‘17

- *You Only Look Once: Unified, Real-Time Object Detection*, Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi ‘15

# Further results and future challenges

# Generative Adversarial Networks - GANs

Source: *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, T. Karras, T. Aila, S. Laine, J. Lehtinen ‘17

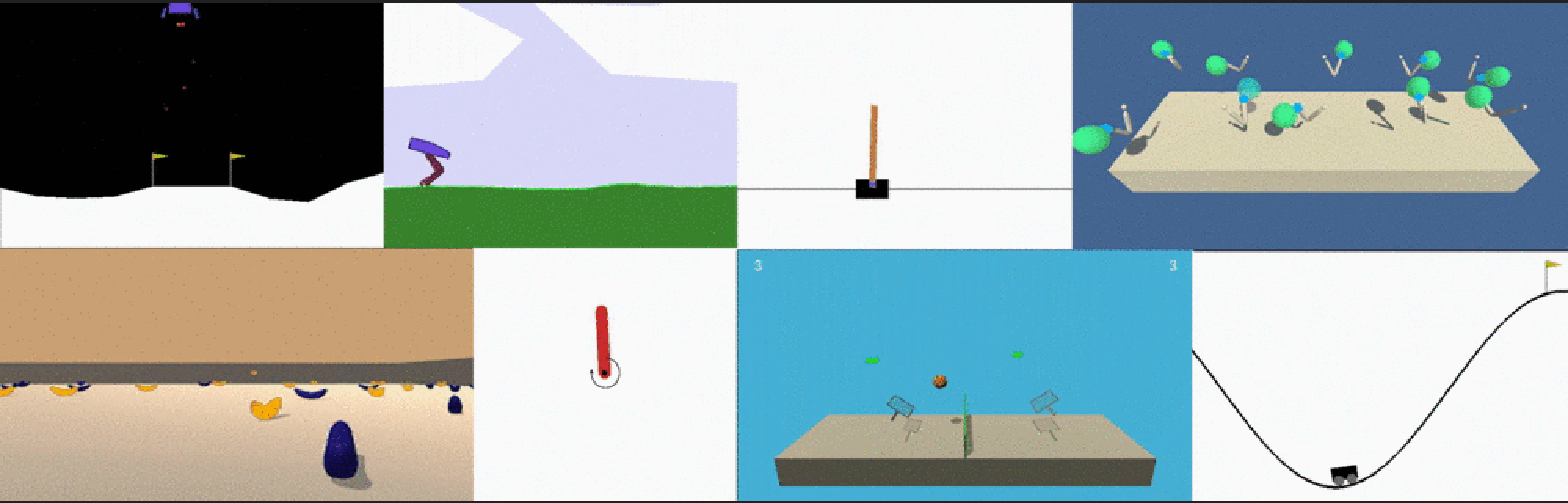
# GAN: ?

One of the following players was generated with a GAN. Which?

# All of them were generated!

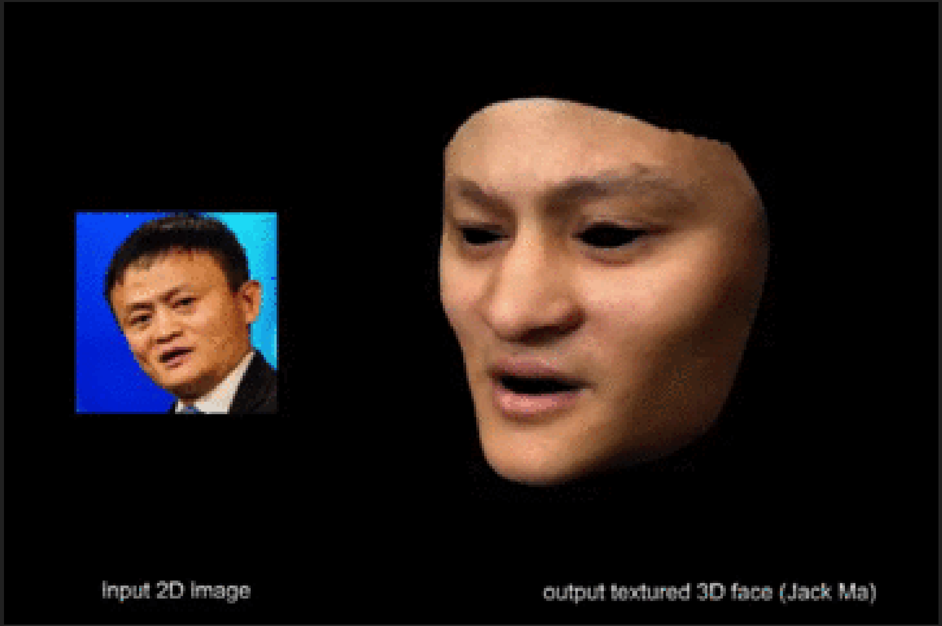
# GAN in Golden Boot

# Deep reinforcement learning




Source: https://medium.com/udacity

# Face model generation




Source: Source: *Photorealistic Facial Texture Inference Using Deep Neural Networks*, S. Saito, L. Wei, L. Hu, K. Nagano, H. Li '16

# Challenges

- Object detection
- Reinforcement learning
- GANs
- …

We want to collaborate!

Contact: michalw@nordeus.com, milosmi@nordeus.com

# Hvala na pažnji!