Intuition on modern deep learning approaches in computer vision

Yaroslav Protsenko
Software engineer at FORTIFIER
Applications of Computer Vision

- Recognition
- Identification
- Movement estimation
- Pose estimation
- 3D reconstruction
Why vision is hard for computers?
High needs for performance

- AlexNet (2012) - 2 GPUs, 1 week
- Clarifai (2013) - 1 GPU, 10 days
- VGG ConvNet (2014) - 2 GPUs, 2-3 weeks
- VGG ConvNet (2016) - 1 GPU, 5 days
Neighbor pixels are similar
Information is encoded in local areas
Optical effects
Transformations
High level of abstraction

So it is a horse

Horse head

Body

Legs

Legs
No formal description of the problem

Human: Where is the flower?

Computer: Which flower?
So why vision is hard for computers?

Computer needs a model of real world, but real world is very complex

- performance
- optical effects
- abstraction
- local areas
- informal problem
- transformations
What is a convolutional neural network?
Convolution
Convolution
Pooling

Max(1, 1, 5, 6) = 6

max pool with 2x2 filters and stride 2
Activation functions
AlexNet (a bit simplified)
Inception architecture

- [https://habrahabr.ru/post/301084/](https://habrahabr.ru/post/301084/)
- [https://habrahabr.ru/post/302242/](https://habrahabr.ru/post/302242/)
- [https://habrahabr.ru/post/303196/](https://habrahabr.ru/post/303196/)
- Search it on Habrahabr by phrase “Эволюция нейросетей в Google”
Vanishing gradient problem

Information → → → → → → → → →

Trainability ← ← ← ← ← ← ← ← ←

Layers

INPUT → → → → → → → → OUTPUT
Vanishing gradient problem

- Other initialization methods
- Use ReLU
- Residual networks
Residual networks

Regular network

Residual network
Thanks!