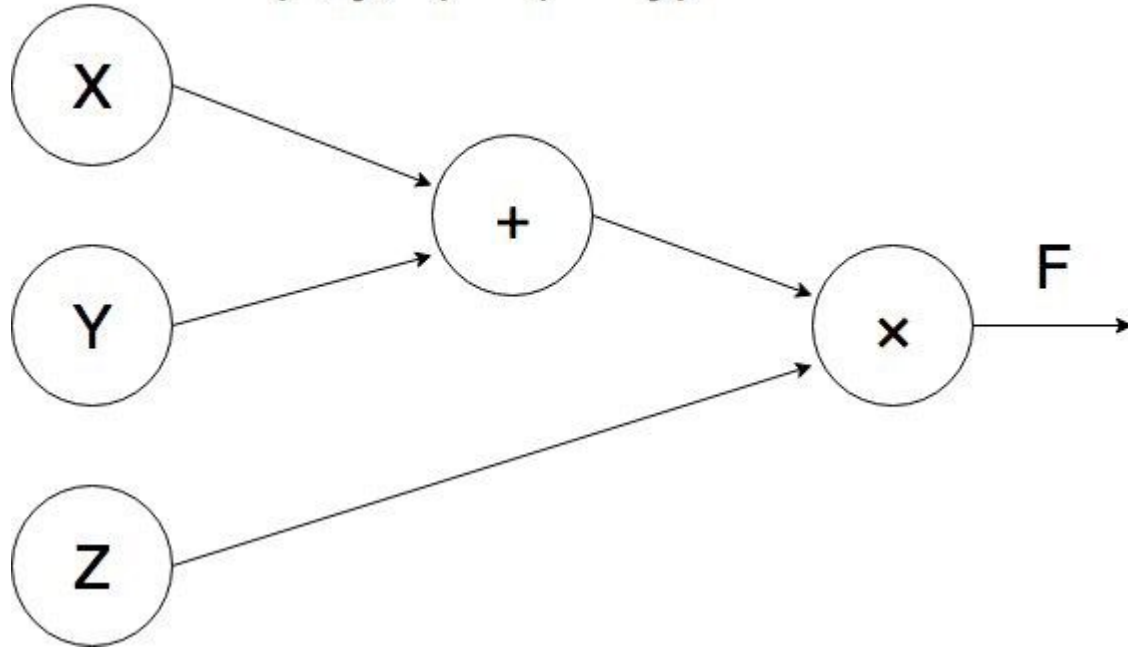


# Training Neural Networks with Local Error Signals

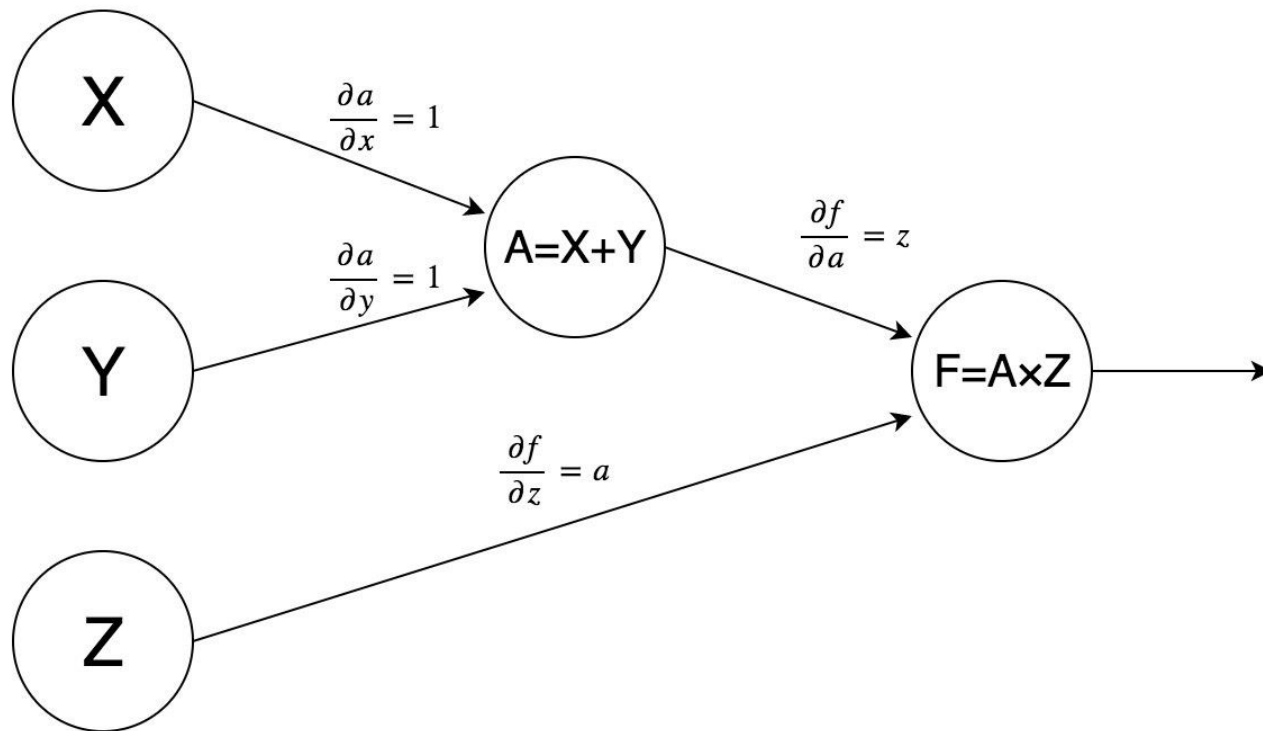
Докладчик: Октай Татанов

# Backpropagation

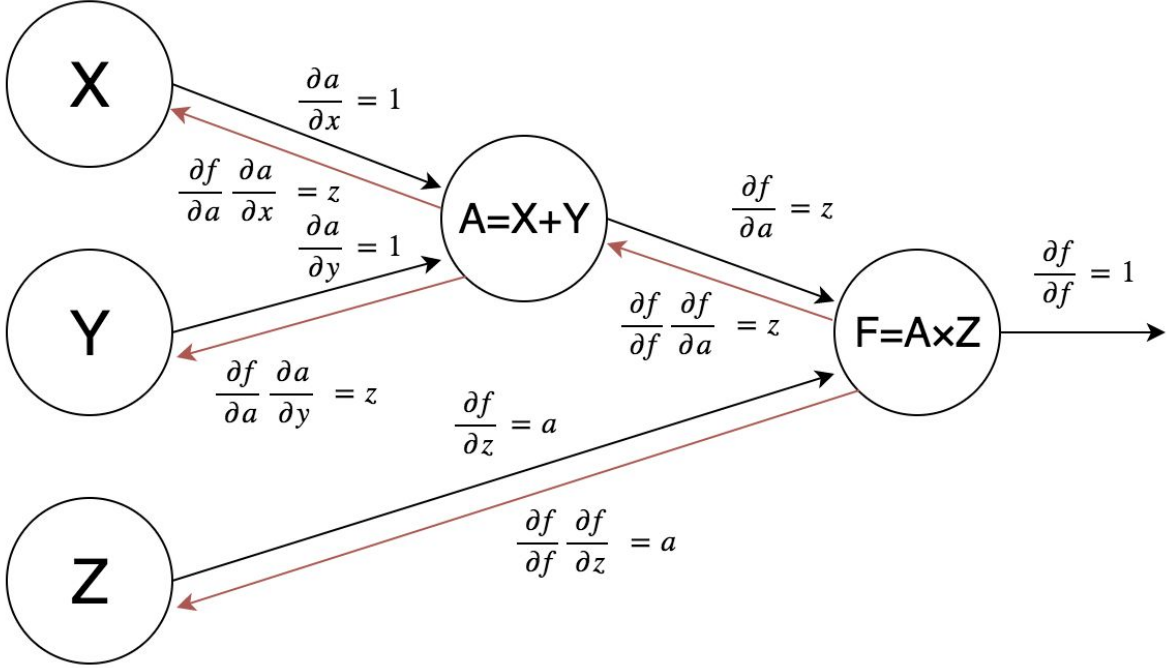
$$F(x, y, z) = (x + y) \times z$$



# Backpropagation: forward



# Backpropagation: backward



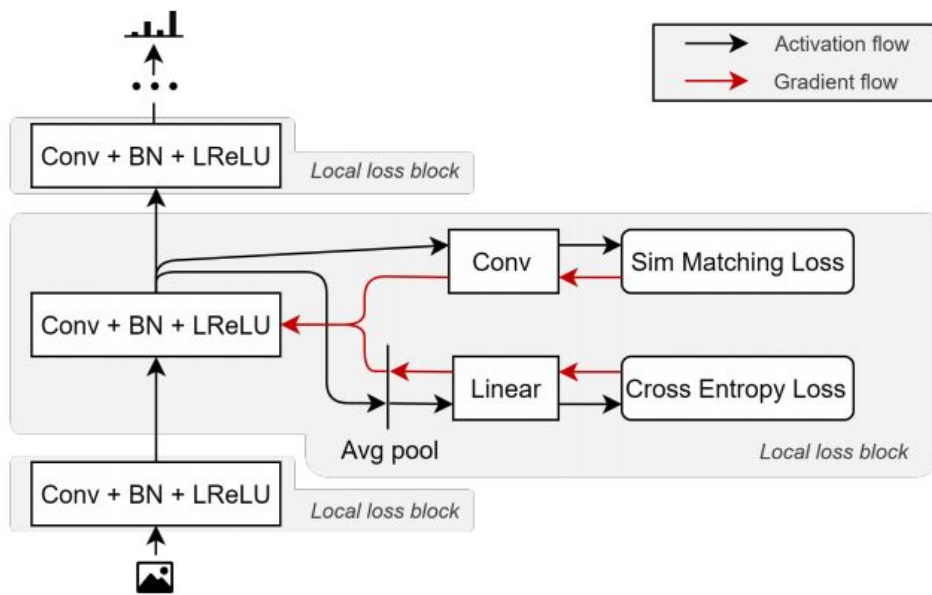
# Backpropagation: cons

Недостатки алгоритма обратного распространения ошибки:

- “backward locking”: невозможность обновить веса модели прежде чем полностью завершится forward и backward
- требует большого количества памяти
- не является биологически правдоподобным

# Alternative approach: layer-wise loss functions

Вместо глобально распространяющихся ошибок, каждый слой обучается **локальным** сигналом, который не распространяется обратно по сети



# Notation

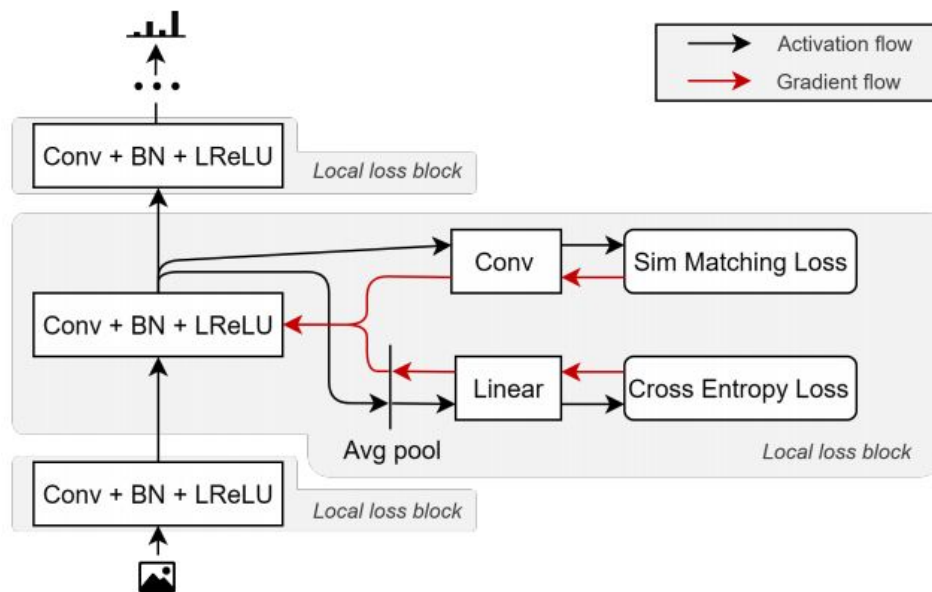
$H = (h_1, \dots, h_n)$  — выход со слоя

$Y = (y_1, \dots, y_n)$  — one-hot лейблы

$S(X)$  — корреляционная матрица:

$$s_{ij} = s_{ji} = \frac{\tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_j}{\|\tilde{\mathbf{x}}_i\|_2 \|\tilde{\mathbf{x}}_j\|_2}$$

$\tilde{\mathbf{x}}_i$  — mean-centered вектор



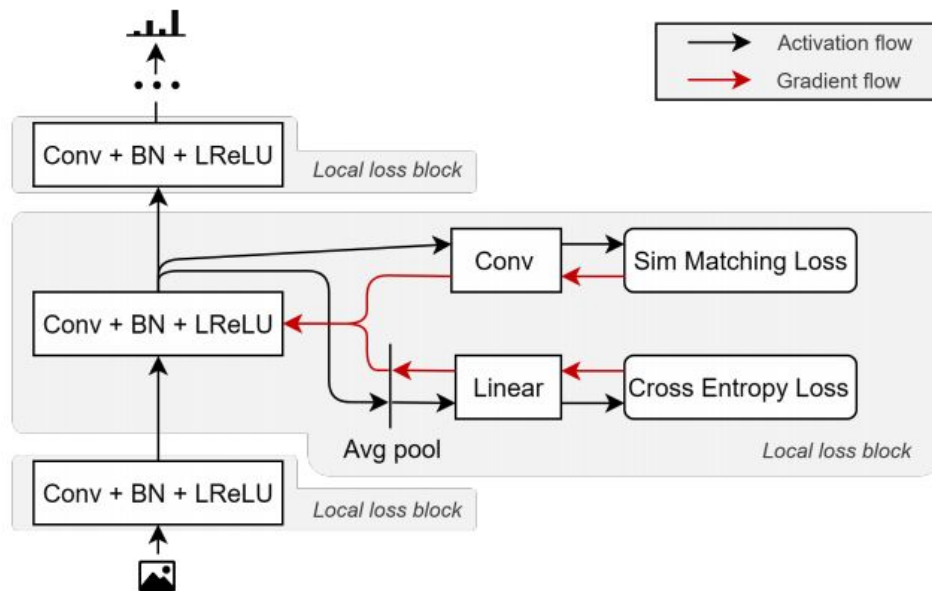
# Prediction (cross-entropy) loss

$$L_{pred} = \text{CrossEntropy}(Y, W^T H)$$

$W$  — веса linear слоя.

Если  $H$  выход из conv слоя:

$H = \text{avg-pooling}(H)$





# Prediction (cross-entropy) loss

$$L_{pred} = \text{CrossEntropy}(Y, W^T H)$$

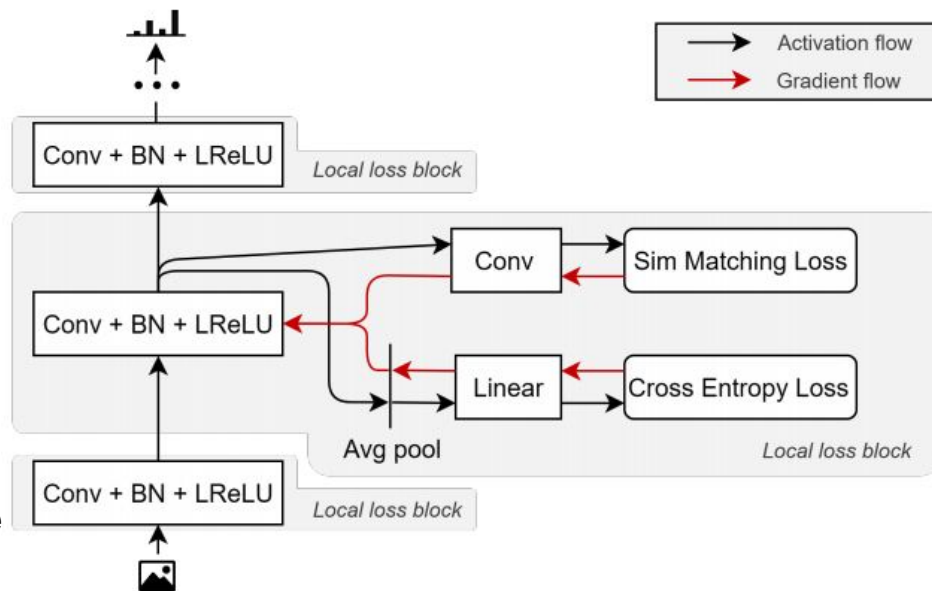
$W$  — веса linear слоя.

Если  $H$  выход из conv слоя:

$H = \text{avg-pooling}(H)$

В сторону биологической правдоподобности:

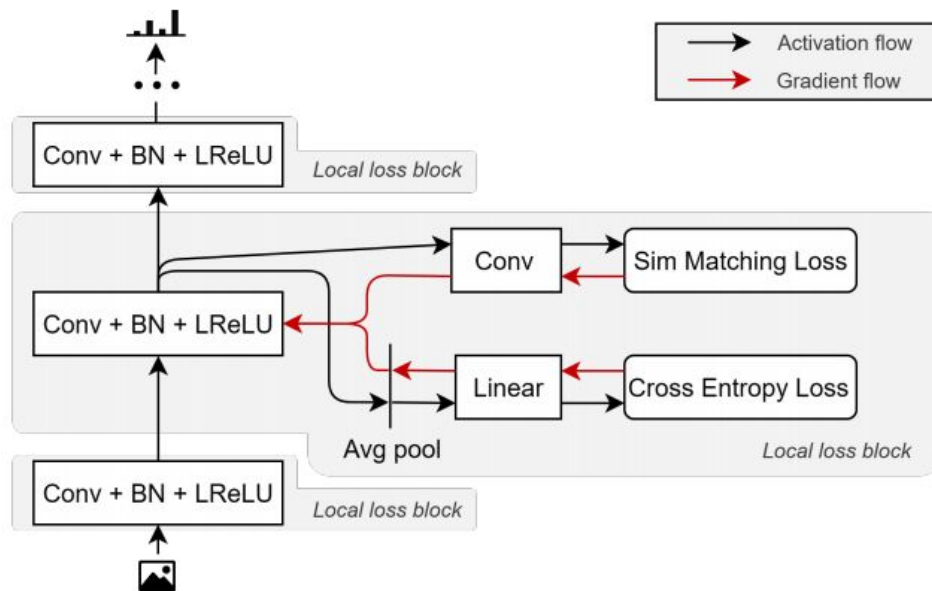
- “**feedback alignment**”: заменяем  $W$  в backward на фиксированную случайную матрицу
- “**binarized random transformation of the target**”: прогоняем исходные лейблы с помощью фиксированного Linear и превращаем в 0/1 по знаку, а обучаем binary cross-entropy



# Similarity matching loss

$$L_{sim} = \|S(NeuralNet(H)) - S(Y)\|_F^2$$

Если  $H$  -- выход из conv слоя:  
NeuralNet = conv + std by feature maps  
Если  $H$  -- выход из linear слоя:  
NeuralNet = linear



# Similarity matching loss

$$L_{sim} = \|S(NeuralNet(H)) - S(Y)\|_F^2$$

Если H -- выход из conv слоя:

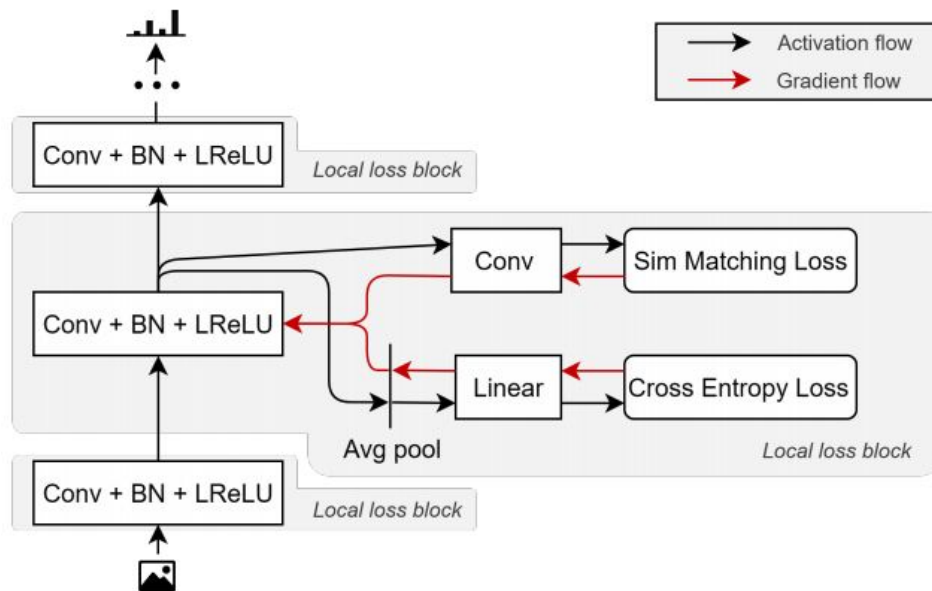
NeuralNet = conv + std by feature maps

Если H -- выход из linear слоя:

NeuralNet = linear

В сторону биологической правдоподобности:

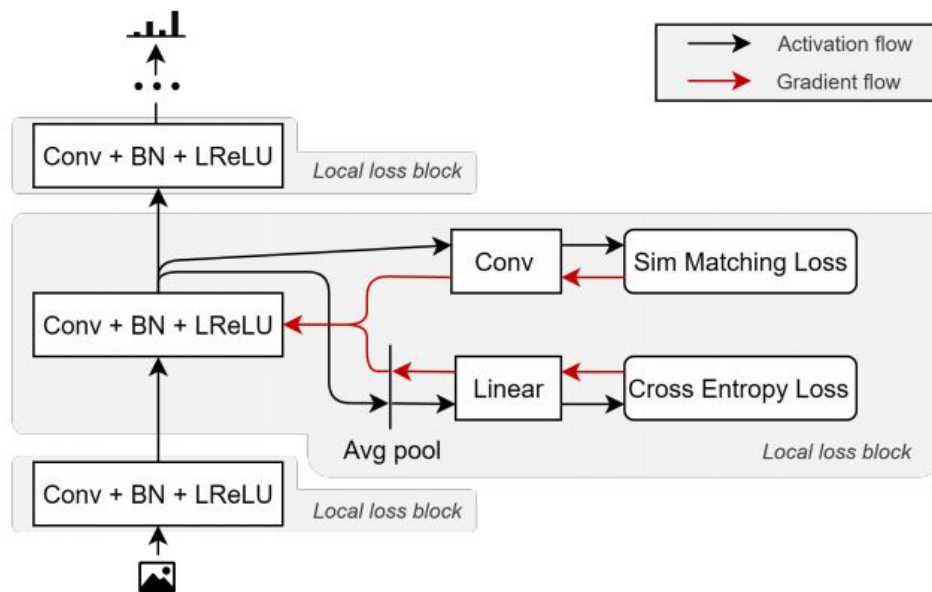
- Убираем conv/linear из NeuralNet
- **“random transformation of the target”**: прогоняем исходные лейблы с помощью фиксированного Linear



# Layer-wise loss

$$L_{predsim} = (1 - \beta)L_{pred} + \beta L_{sim}$$

В статье использовали  $\beta = 0.99$ .



# Experiments & discussion

- prediction loss достигает довольно близких значений к backprop
- similarity loss позволяет обучать модели
- pred+sim вместе работают хорошо
- отличные результаты для VGG-like архитектур, не очень для ResNet-like

Table 4. CIFAR10 with standard data augmentation. Test error in percent.

Model	#par	glob	Local loss functions		
			pred	sim	predsim
3x3000 MLP	27M	33.6	32.3	33.5	<b>30.9</b>
VGG8B	8.9M	5.99	8.40	7.16	<b>5.58</b>
VGG11B	12M	5.56	8.39	6.70	<b>5.30</b>
VGG11B(2x)	42M	4.91	7.30	6.66	<b>4.42</b>
VGG11B(3x)	91M	5.02	7.37	9.34 <sup>3</sup>	<b>3.97</b>
11B(3x)+CO	91M	-	-	-	3.60
WRN	56M	3.87	-	-	-
WRN+CO	56M	3.08	-	-	-

Table 6. CIFAR100 with standard data augmentation. Test error in percent.

Model	#par	glob	Local loss functions		
			pred	sim	predsim
3x3000 MLP	27M	62.6	58.9	62.5	<b>56.9</b>
VGG8B	9.0M	26.2	29.3	32.6	<b>24.1</b>
VGG11B	12M	25.2	29.6	30.8	<b>24.1</b>
VGG11B(2x)	42M	23.4	26.9	28.0	<b>21.2</b>
VGG11B(3x)	91M	23.7	25.9	28.0	<b>20.1</b>
WRN	56M	18.8	-	-	-
WRN+CO	56M	18.4	-	-	-

# Experiments & discussion

Предложенный подход не только решает многие проблемы backprop, а так же неплохо работает и в связке с ним

*Table 9. Similarity matching as a complementary objective. Test error in percent.*

Dataset	Model	#par	glob	predsim	glob+sim
MNIST	VGG8B	7.3M	0.26	0.31	<b>0.24</b>
Fashion-MNIST	VGG8B	7.3M	4.53	4.65	<b>4.16</b>
Kuzushiji-MNIST	VGG8B	7.3M	1.53	1.36	<b>1.13</b>
CIFAR-10	VGG11B	12M	5.56	5.30	<b>4.33</b>
CIFAR-100	VGG11B	12M	25.2	24.1	<b>22.2</b>
SVHN	VGG8B	8.9M	2.29	<b>1.74</b>	1.95
STL-10	VGG8B	12M	33.1	<b>20.5</b>	25.6

# Experiments & discussion

Layer-wise loss сам по себе является сильным регуляризатором

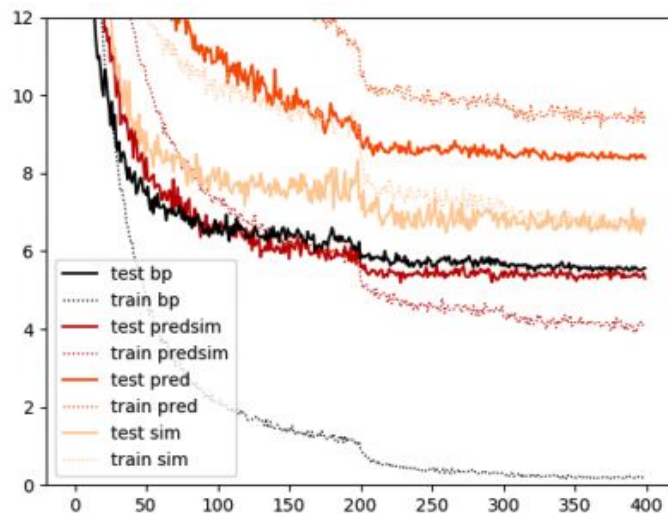


Figure 2. Training and test classification errors on CIFAR10 with full backprop and different local loss functions, with a VGG11B(1x) network.

# Experiments & discussion

Layer-wise loss, измененный в сторону биологической правдоподобности, дает лучшие результаты на CIFAR-10 среди layer-wise loss без backprop

*Table 5. CIFAR10 with standard data augmentation. No back-propagation. Test error in percent.*

Model	#par	pred-bio	sim-bio	predsim-bio
VGG8B	8.9M	9.80	13.39	<b>9.02</b>
VGG8B(2x)	31M	-	-	7.80