

# Генеративные модели в SSL для разных типов задач

- Цесарев Даниил Игоревич
- Лебедюк Вероника Владимировна
  - Марьин Геннадий Аркадьевич
  - Сечкарь Константин Сергеевич

# Краткий экскурс

**Цель:** Установить насколько универсален новый метод - использование генеративной модели - в контексте SSL

Method	Evaluation	Params (M)	Acc. %
<i>on CIFAR-10</i>			
WideResNet-28-10 [63]	Supervised	36	96.3
<b>DDAE (EDM)</b>	Linear	36	95.9
SimCLR Res-50 [11]	Linear	24	94.0
SimCLRv2 Res-101-SK [12]	Linear	65	96.4
<b>DDAE (EDM)</b>	Fine-tune	36	97.2
SimCLRv2 Res-101-SK [12]	Fine-tune	65	97.1

**Задачи, которые перед нами стояли:**

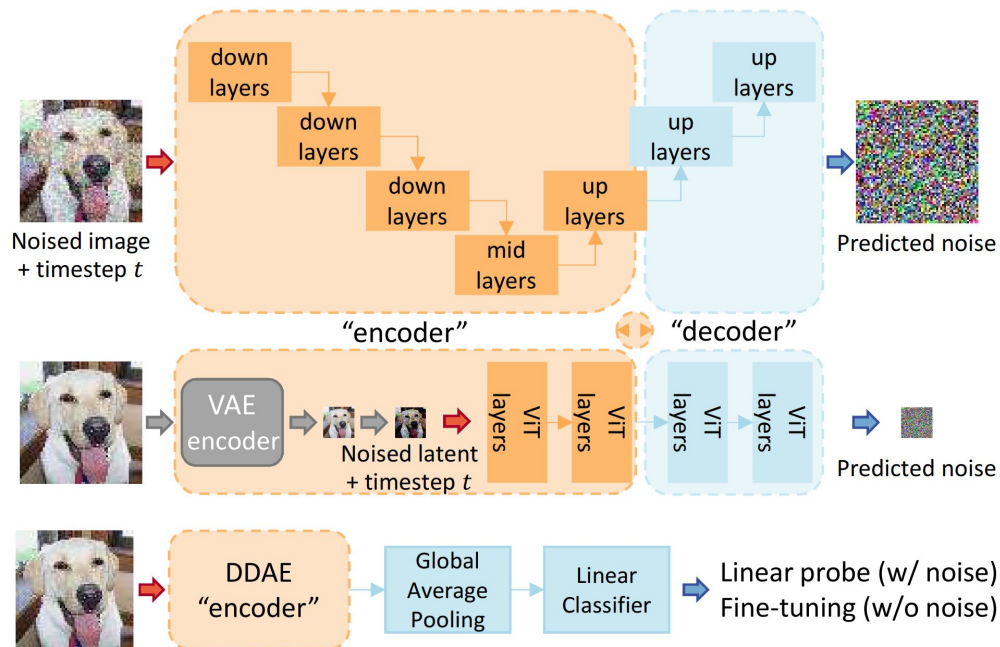
- Реализовать диффузионный энкодер с архитектурой подходящей для любой конечной задачи
- Сравнить качество получающихся репрезентаций с SOTA бейзлайнами
- Поставить дальнейшие эксперименты для оптимальности энкодера

DDAE -  
Классификация

# DDAE - обзор

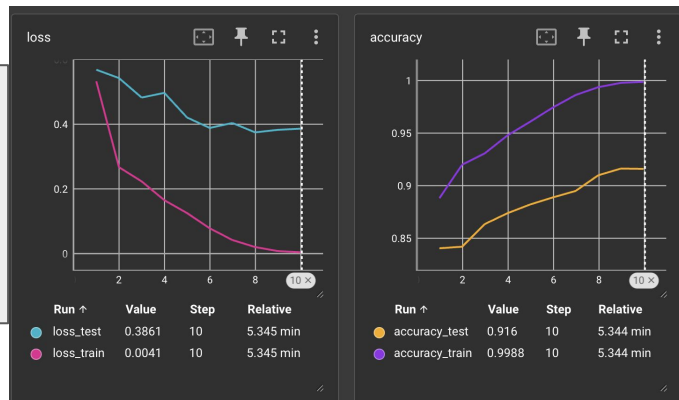
**Идея:** решая задачу безусловной генерации модель формирует хорошие знания об изображениях

Method	Evaluation	Params (M)	Acc. %
<i>on CIFAR-10</i>			
WideResNet-28-10 [63]	Supervised	36	96.3
<b>DDAE (EDM)</b>	Linear	36	95.9
SimCLR Res-50 [11]	Linear	24	94.0
SimCLRv2 Res-101-SK [12]	Linear	65	96.4
<b>DDAE (EDM)</b>	Fine-tune	36	97.2
SimCLRv2 Res-101-SK [12]	Fine-tune	65	97.1



# DDAE - классификация

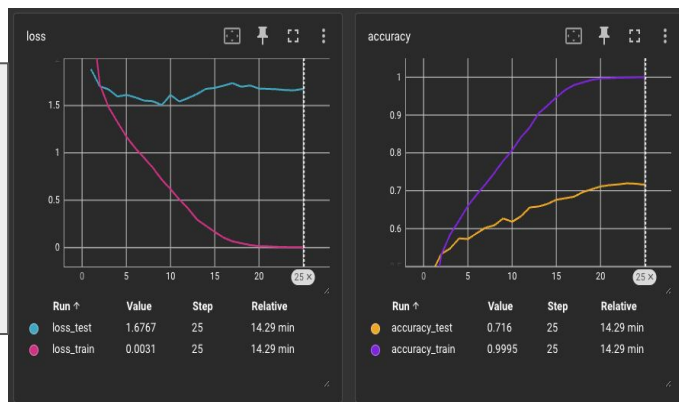
CIFAR10



## Результаты:

- CIFAR10 accuracy = 0.916
- CIFAR100 accuracy = 0.716

CIFAR100



**Итог:** результаты статьи не воспроизводились, однако метрики оказались приемлемы

# DDAE - Best Practice on CIFAR10

## Шум

	Accuracy
No noize	0.908
Noize	0.913
ts50	0.907
Ema, ts50	0.912
Ema, ts20	0.913
Ema, ts10	<b>0.916</b>

## Слой репрезентаций + обратка выхода

	Accuracy
Слой 4/4	0.905
Слой 2/4	<u>0.905</u>
Слой 1/4	0.903
Слой 2/4 + свертка вместо пулинга	<b>0.907</b>

## Оптимайзер

	Accuracy
Adam	0.907
AdamW	<b>0.913</b>

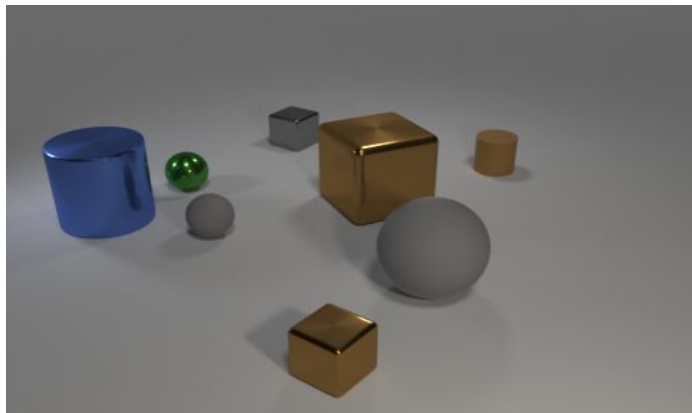
## Note CIFAR100:

	Accuracy
No noize	0.55
Noize ts50	<b>0.716</b>

SimCLR  
Object Count

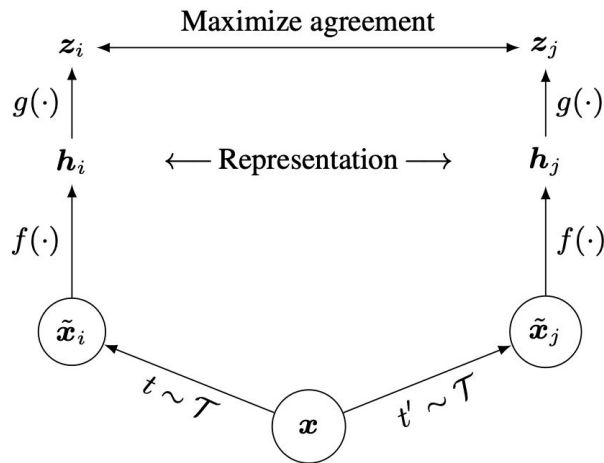
# Подсчет количества объектов на изображении

**CLEVR** - датасет для ответов на вопросы по изображению



**Кол-во** объектов на изображениях от 3 до 7

Train	70k
Test	15k



**SimCLR** - классический контрастный подход.

Для задачи выбраны аугментации:

- Random Resized Crop + Horizontal Flip
- Color Jitter
- Random GrayScale



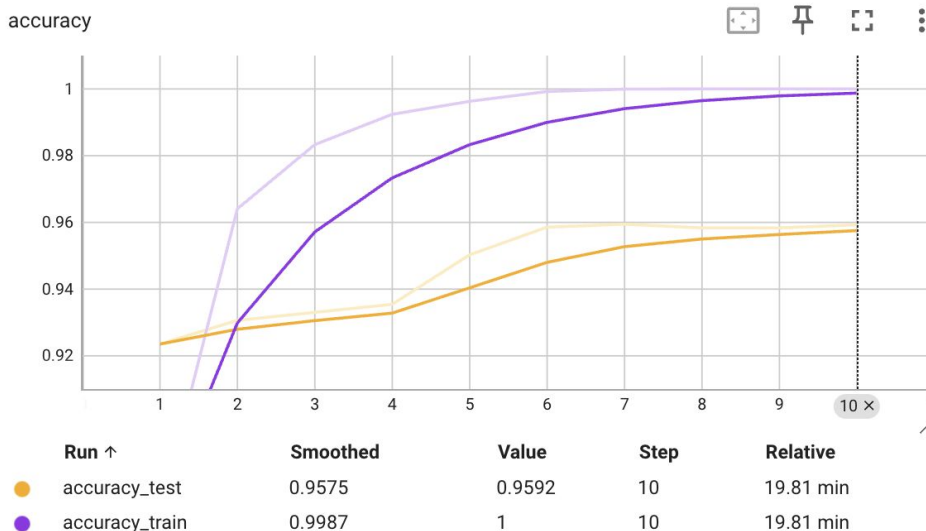
# SimCLR baseline

Тип  
задачи:  
Semantic  
Pixelwise

Датасет:  
CLEVR

Бейзлайн:  
SimCLR

## Сравнение на CIFAR10



## Object Count

CLEVR	Accuracy
SimCLR	<b>0.55</b>
Diffusion	0.46

# Depth prediction

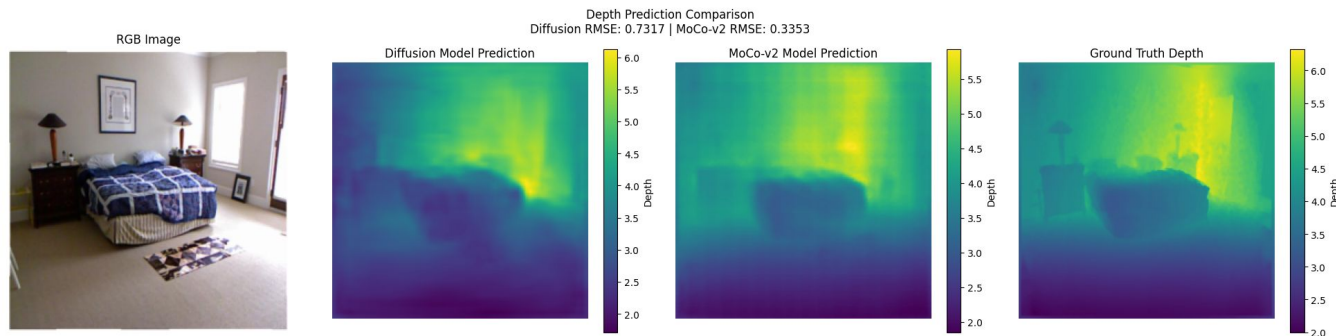
# Depth prediction

**Тип задачи:**  
Structural  
Pixelwise

**Датасет:**  
NYU Depth V2

**Бейзлайн:**  
MoCov2

**End-task  
архитектура:**  
U-Net



**Бейзлайн**

Ключевые особенности:

- Momentum encoder
- Dynamic queue

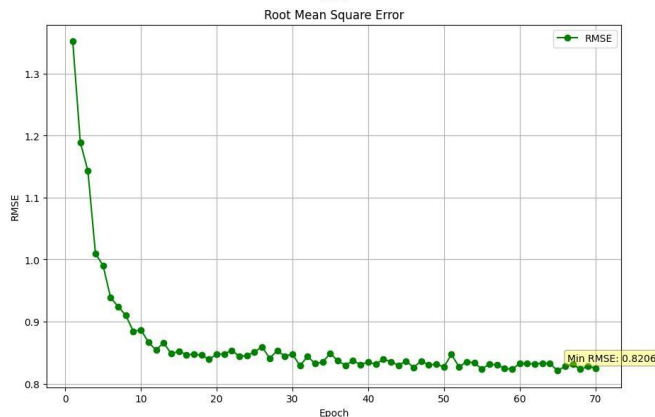
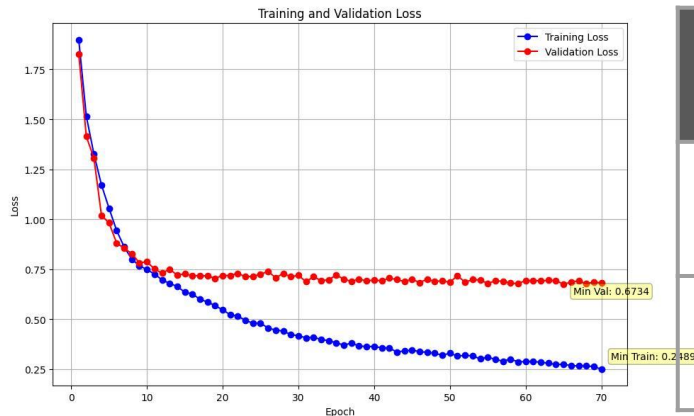
# Depth prediction

Тип задачи:  
Structural  
Pixelwise

Датасет:  
NYU Depth V2

Бейзлайн:  
MoCov2

End-task  
архитектура:  
U-Net



	RMSE
MoCov2	<b>0.69</b>
Diffusion	0.82

	L1-Loss
MoCov2	<b>0.47</b>
Diffusion	0.68

# Сегментация

# Сегментация

Тип задачи:  
Semantic  
Pixelwise

Датасет:  
Oxford-IIIT Pet

Бейзлайн:  
SwAV

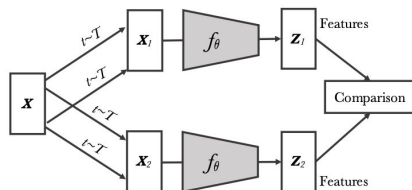
End-task  
архитектура:  
DeepLabV3+

## Бейзлайн

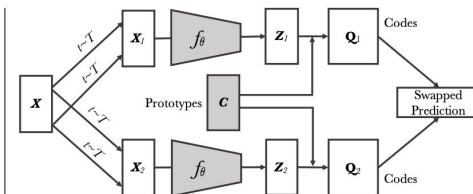
SwAV avoids explicit pairwise feature comparisons, contrasting between multiple image views by comparing their cluster assignments

$z_t, z_s$  image features from two different augmentations of the same image

$q_t, q_s$  their codes computed by matching to a set of  $K$  prototypes  $\{c_1 \dots c_K\}$



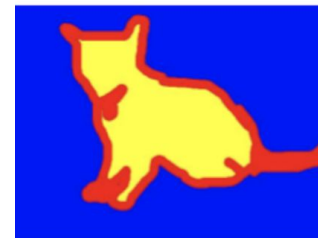
Contrastive instance learning



Swapping Assignments between Views (Ours)

## Датасет:

	trainval	test
n_imgs	3680	3669



# Сегментация

**Тип задачи:**

Semantic  
Pixelwise

**Датасет:**

Oxford-IIIT Pet

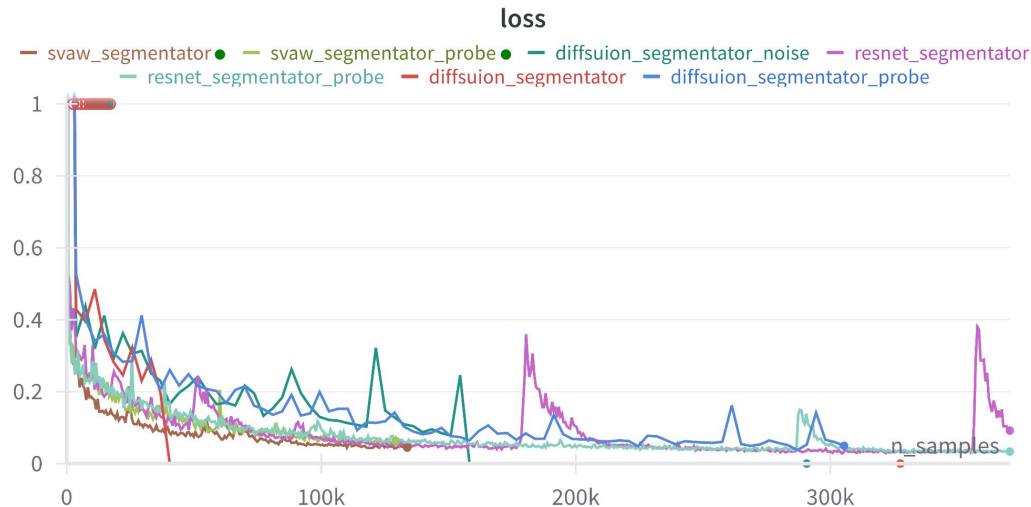
**Бейзлайн:**

SwAV

**End-task**

**архитектура:**

DeepLabV3+



Trainable  
params

	all	probe
resnet	78,773,493	55,265,461
diffusion	140,627,128	26,953,909

# Сегментация

Тип задачи:

Semantic  
Pixelwise

Датасет:

Oxford-IIIT Pet

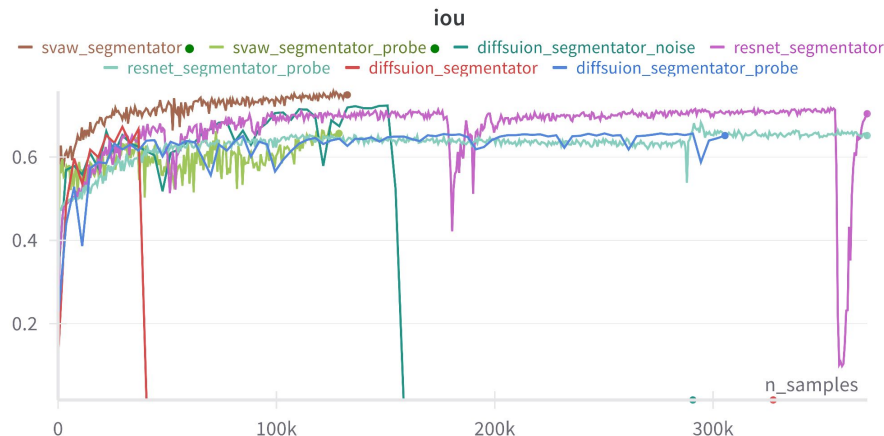
Бейзлайн:

SwAV

End-task

архитектура:

DeepLabV3+

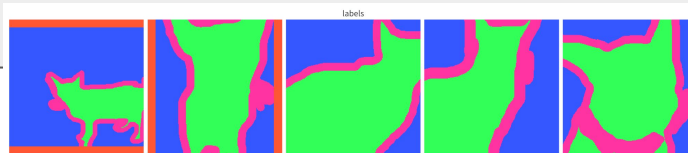


IOU

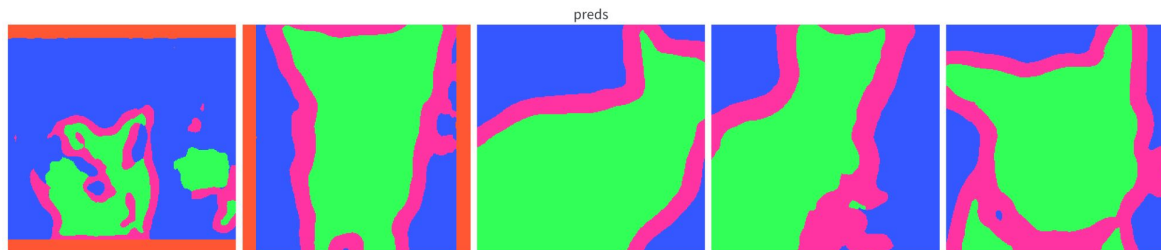
	default	w/ probe	w/ noise
MoCov2	0.7	0.65	-
SwAV	<b>0.74</b>	0.65	-
Diffusion	0.66	0.65	<b>0.72</b>



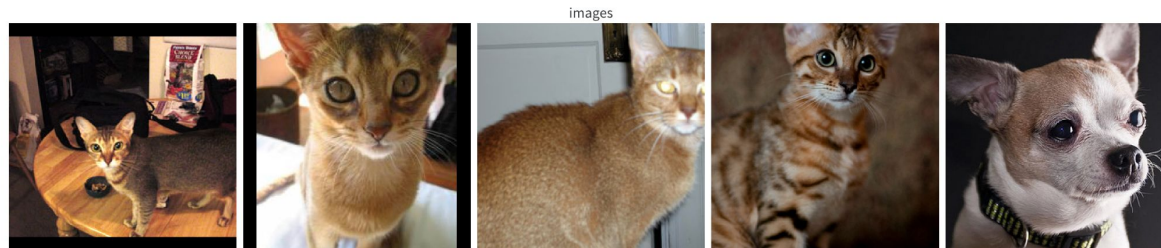
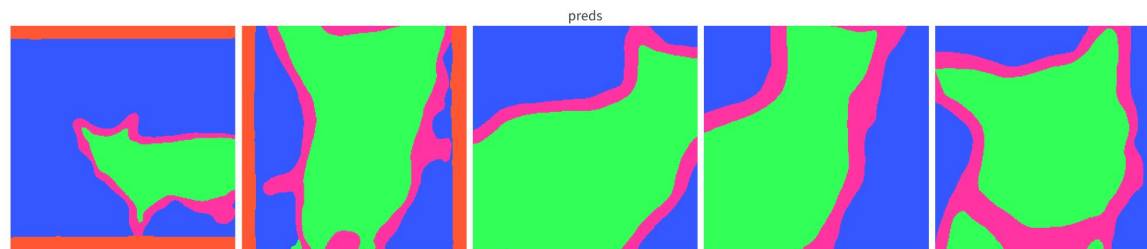
# Сегментация



Diff  
best



resnet  
best  
(swav)



Спасибо за внимание!