Self-Supervised Fine-Grained Food Recognition

Petia Radeva
Universitat de Barcelona &
Computer Vision Center, Spain

UNIVERSITAT DE BARCELONA

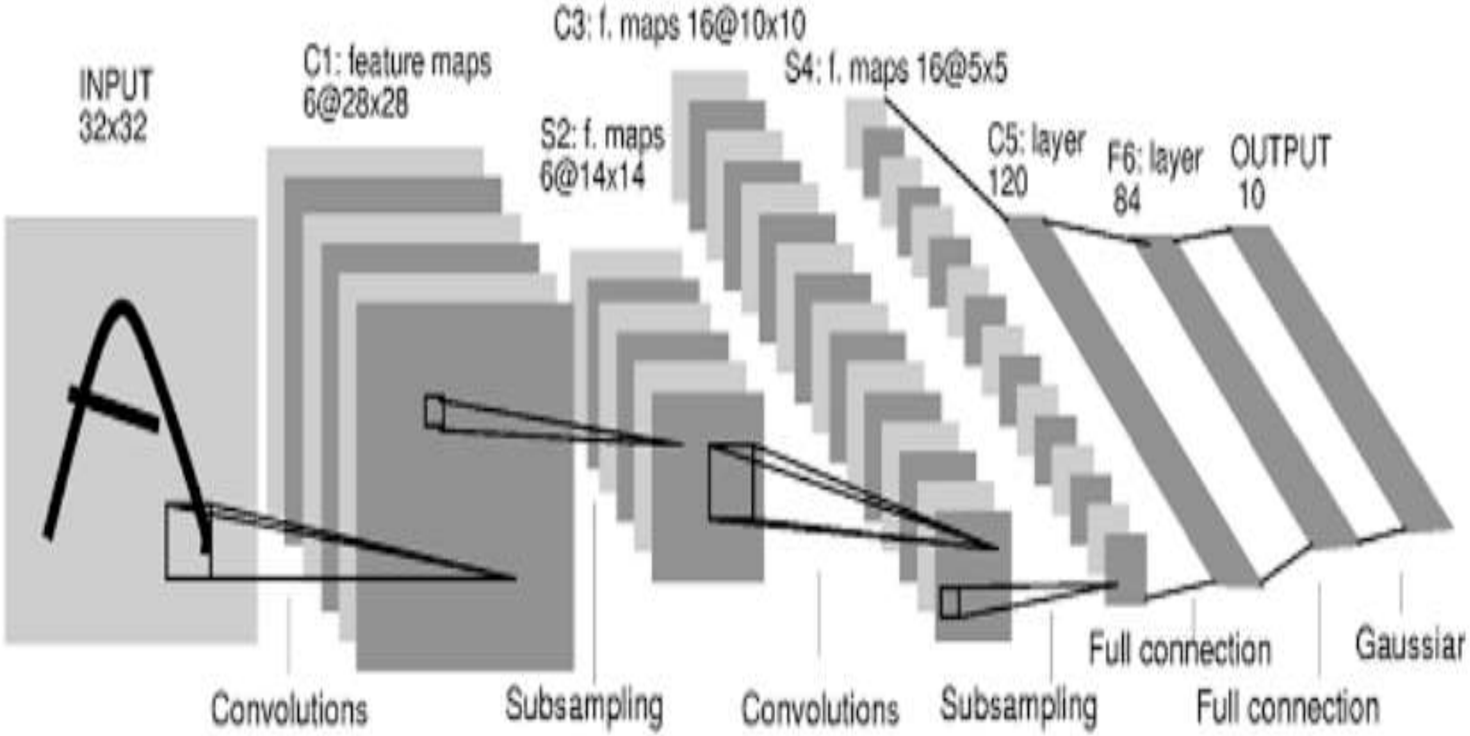
CVC

ICPRAM 2023
12th International Conference on Pattern Recognition Applications and Methods

Which is the year of CNNs?

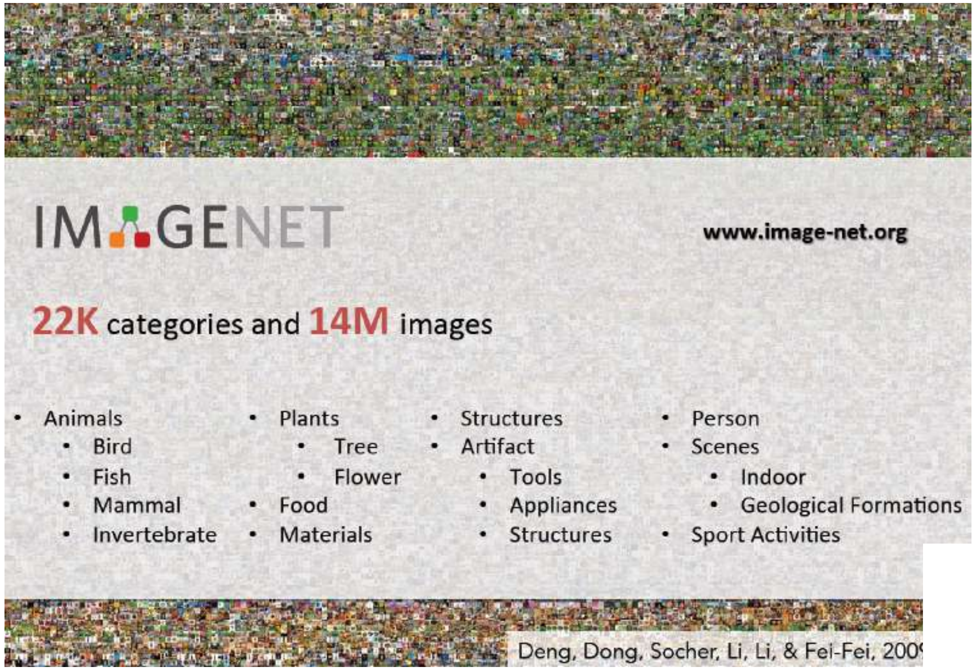
1998

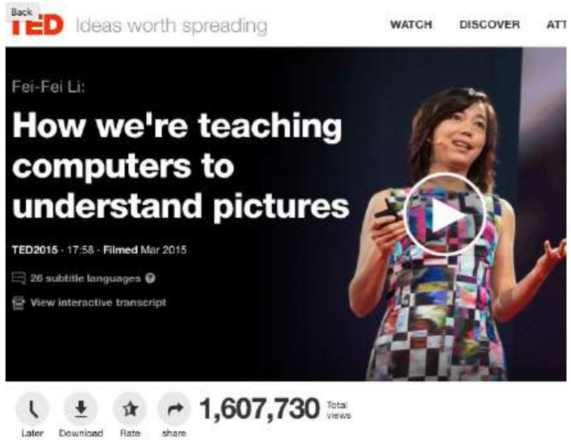
LeCun et al.

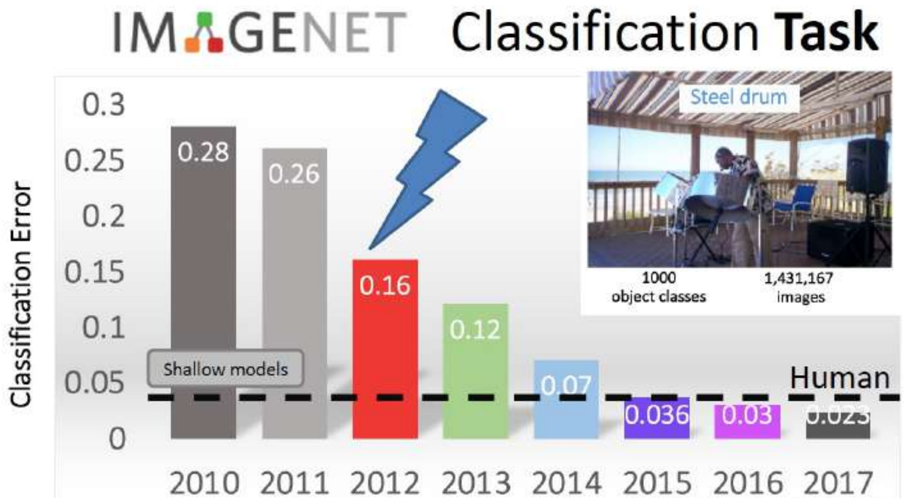


LeCun, Yann; Léon Bottou; Yoshua Bengio; Patrick Haffner (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE* **86** (11): 2278–2324

Imagenet

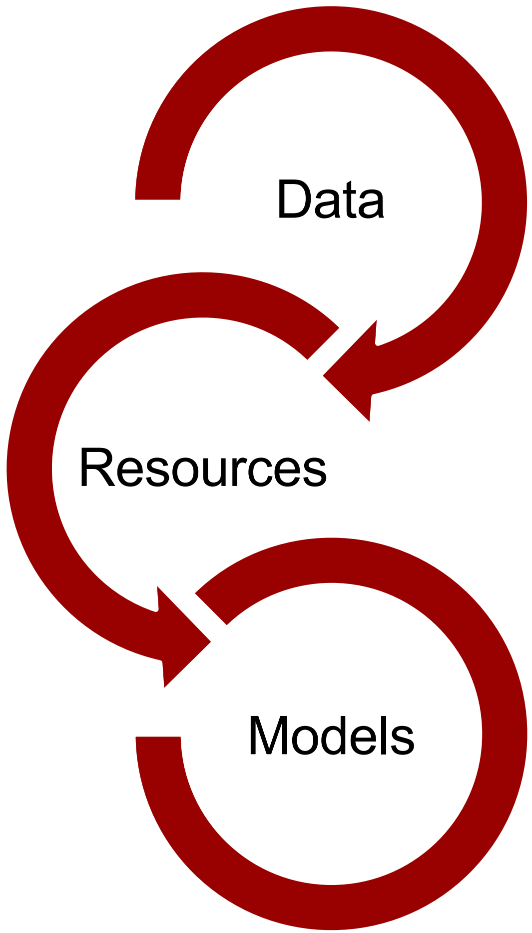


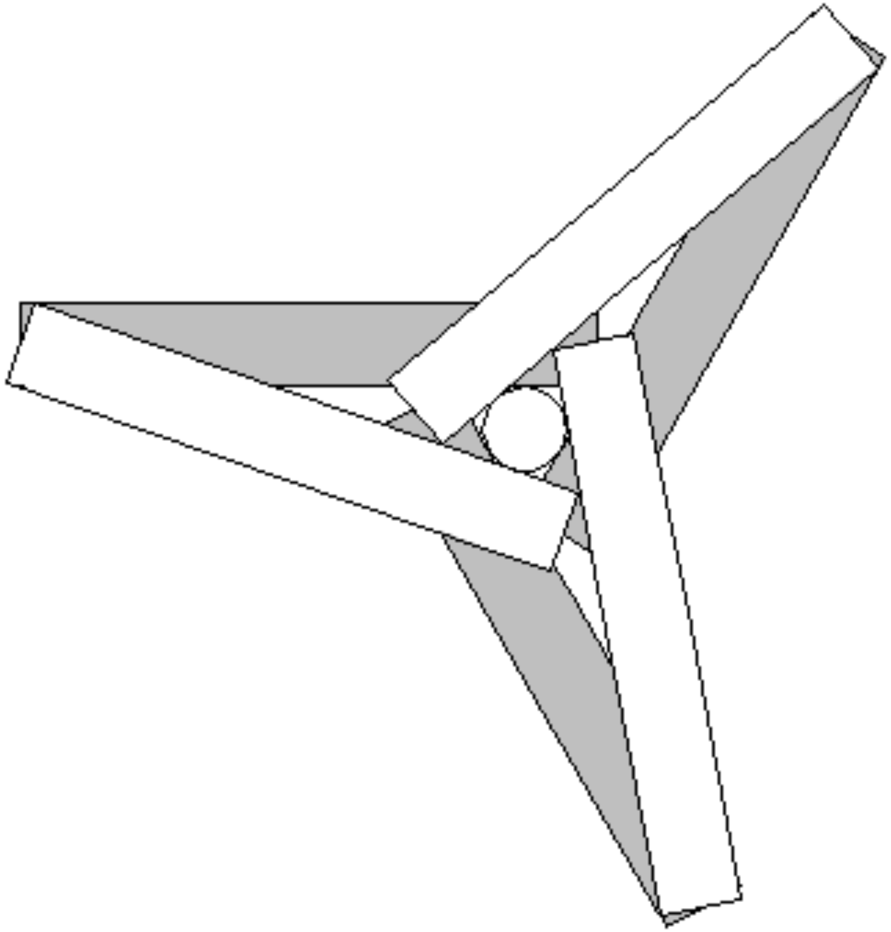




Deng et al. *CVPR*, 2009; Russakovsky et al. *IJCV*, 2012;

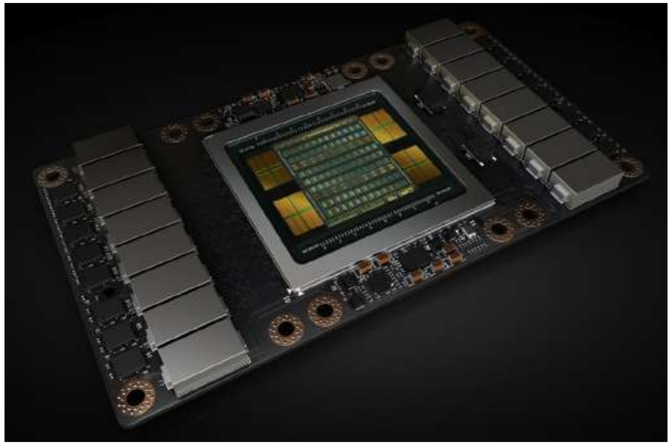
The magic triangle

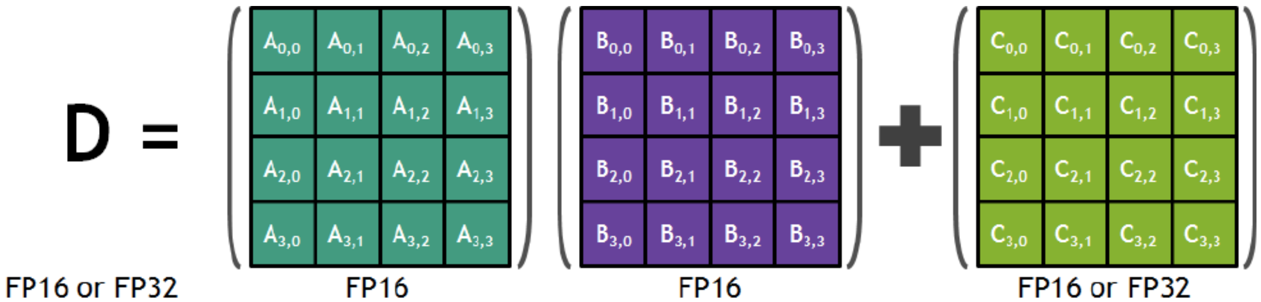




The Importance of GPUs

- Nvidia Tensor Cores - 2017
- Google Tensor Processing Unit (TPU) - 2016
- Intel - Nervana Neural Processor - 2017
- GPUs in Cloud Computing (Google, 2017)

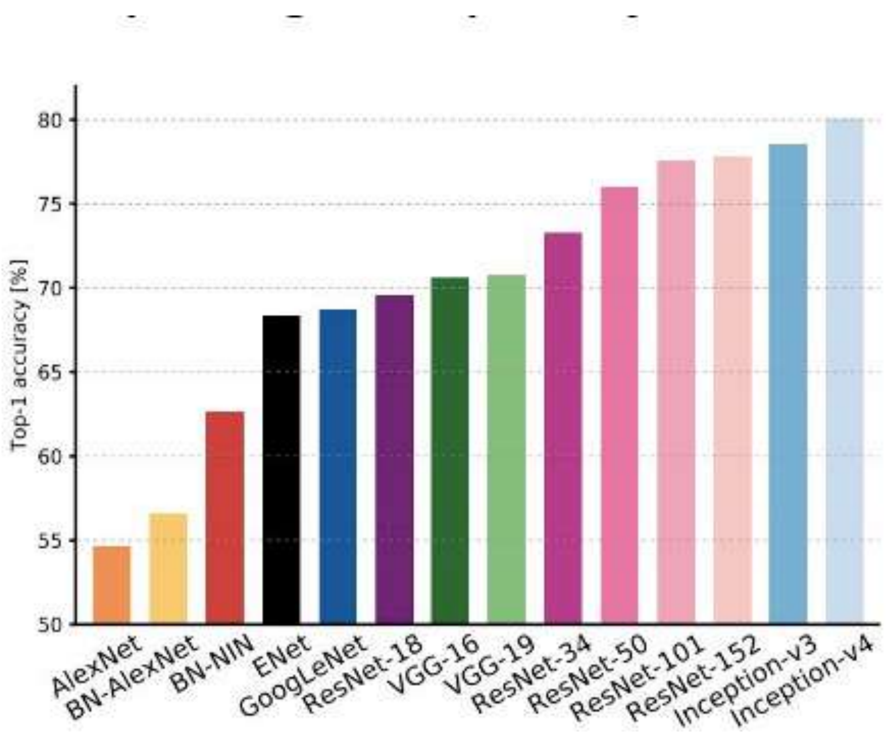


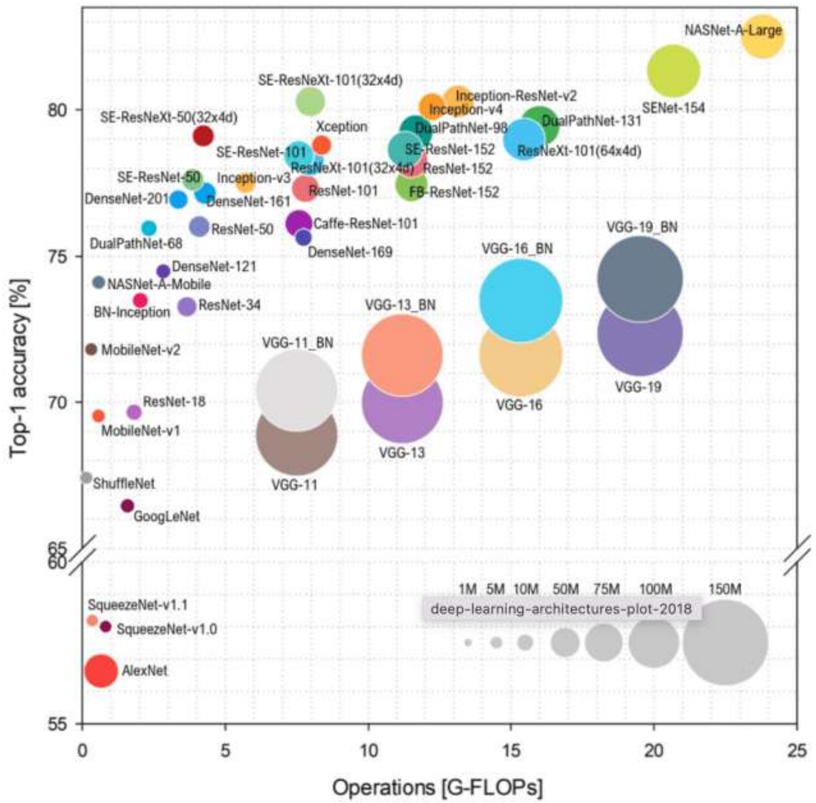


GPU cores is based on matrix multiplication

https://www.doc.ic.ac.uk/~jce317/history-machine-learning.html#top

Available NNs



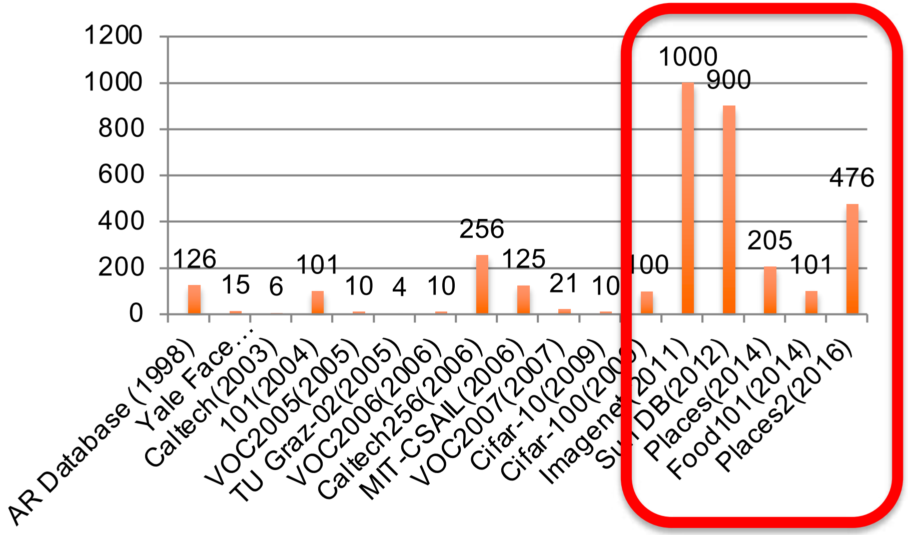


- **Millions of parameters!!!**

The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

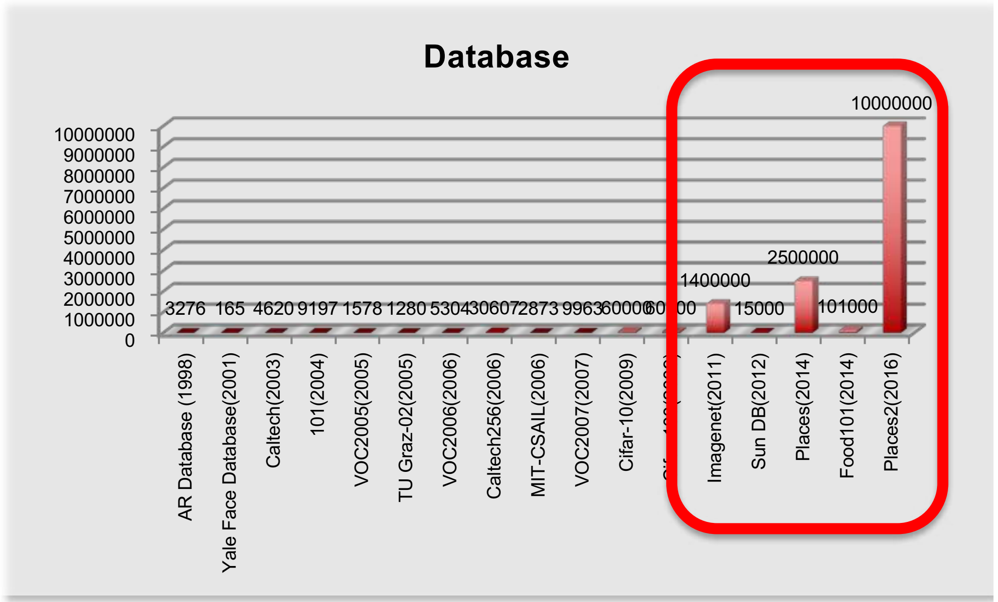
Image databases evolution

Number of objects/Database



ImageNet & Deep learning

Number of images/Database



What are the most popular datasets today?

Dataset	Papers	Benchmarks	Images (K)	Classes	Sizes
Cifar-10	10581	66	60	10	32x32
ImageNet	10046	97	**1400**	**1000**	variable
COCO	7160	78	123	80	
MNIST	5911	49	60	10	28x28
Cifar-100	5322	42	60	100	32x32
Cityskapes	2562	37	25	8	
SVHN	2474	11	60	10	32x32
Kitti	2453	120	0,5	11	
CelebA	2408	20	202	10177	178x218
Fashion-MNIST	2150	17	70	10	28x28
CUB-00-2011	2408	37	12	200	
Places	760	4	**2500**	205	
Tiny ImageNet	516	7	31	200	
Places205	468	1	**2500**	205	
Caltech-101	393	6	5	101	300x200
Stanford Cars	392	8	16	196	360x240
Caltech-256	345	4	30	257	

Large Scale Food Recognition Dataset


Journals & Magazines > IEEE Transactions on Pattern ... > Early Access

Large Scale Visual Food Recognition

Publisher: IEEE | Cite This | PDF

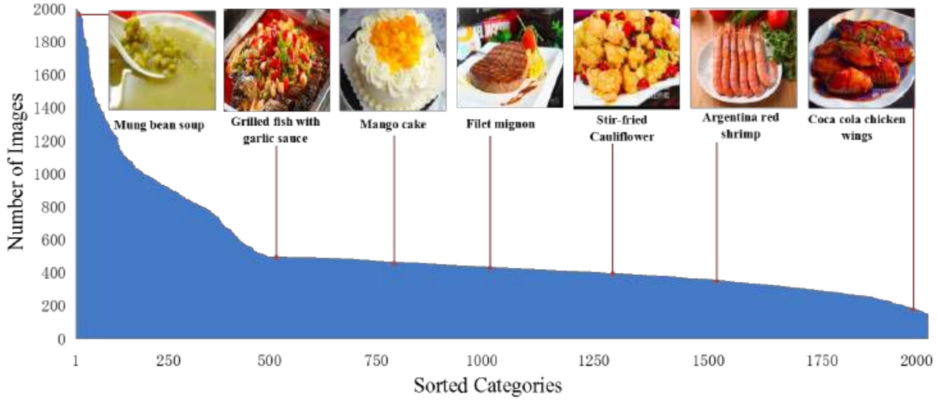
Weiqing Min ; Zhiling Wang ; Yuxin Liu ; Mengjiang Luo ; Liping Kang ; Xiaoming Wei ; Xia... All Authors

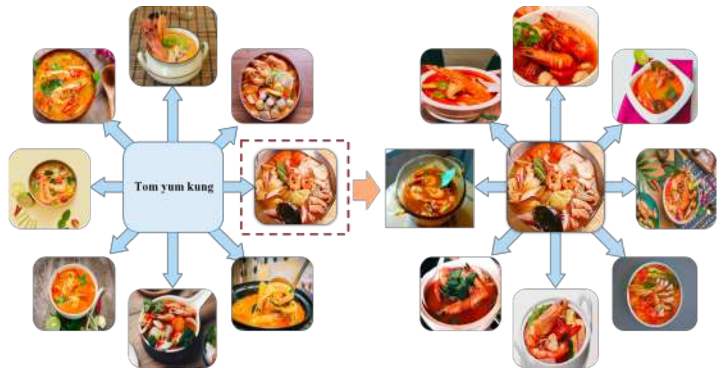
140 Full Text Views

Abstract | Authors | Keywords | Metrics

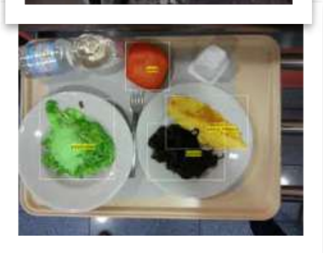
Abstract:
Food recognition plays an important role in food choice and intake, which is essential to the health and well-being of humans. It is thus of importance to the computer vision community, and can further support many food-oriented vision and multimodal tasks, e.g., food detection and segmentation, cross-modal recipe retrieval and generation.

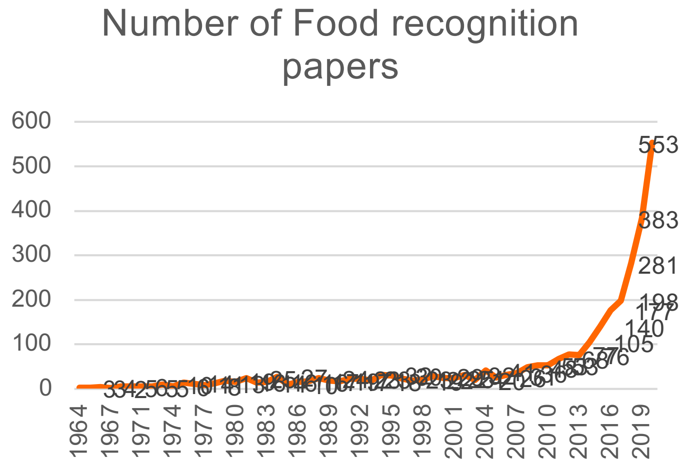
Meat	Cheese back ribs	Tomahaw	Fried pork in scoop	Sheep roll
Vegetables	Eggplant salad	Fruit salad	Shredded cucumber	Fried eggplant
Bread	Tuna pizza	Beef burger	Seafood pancake	Coconut bread
Snack	Egg tart	Roti prata	Strawberry smoothie	Takoyaki
Fried food	Tonkatsu	Fried chicken	Fried cuttlefish balls	Fried tofu
Seafood	Tempura	Spicy crab	Geoduck sashimi	Cod fish steak
Cereal products	Egg fried rice	Salmon sushi	Pan-fried pork bun	Instant noodles



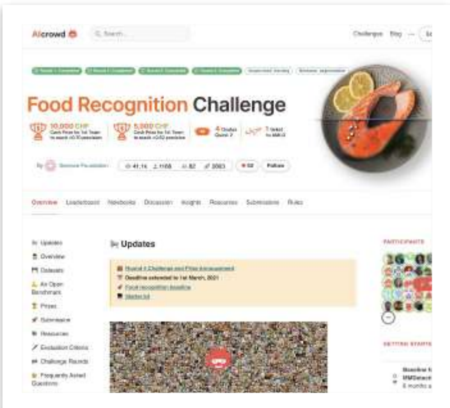


Food recognition popularity





iFood 2011 fine-grained (prepared) food categories with 135733

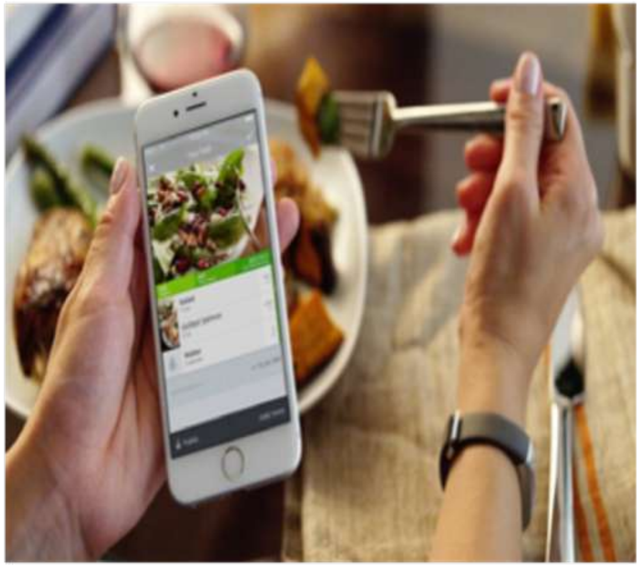


AICrowd: 26000 annotated segmented images

LargeFineFoodAI: 1,000 fine-grained food categories and over 50,000 images.


Food image analysis

Why food recognition?



"Camera eats first"

180M #food

90/minute

54% take picture

39% post it

Why is the food recognition a challenge?

Motivation

Food Analysis Problems

Ingredients

- Intra-class variability
- Inter-class similarity

Decreasing in Precision

Intra-class variability example: Apple. Image source: Recipes5k

Inter-class similarity example: Tomato sauce and Curry sauce. Image source: Recipes5k

The food recognition is a Fine-grained recognition problem

Challenges of Food image analysis

Food256: 25.600 images (100 images/class) Classes: 256

Food101 – 101.000 images
(1000 images/class)
Classes: 101

FoodX-251
Classes: 251
140K images

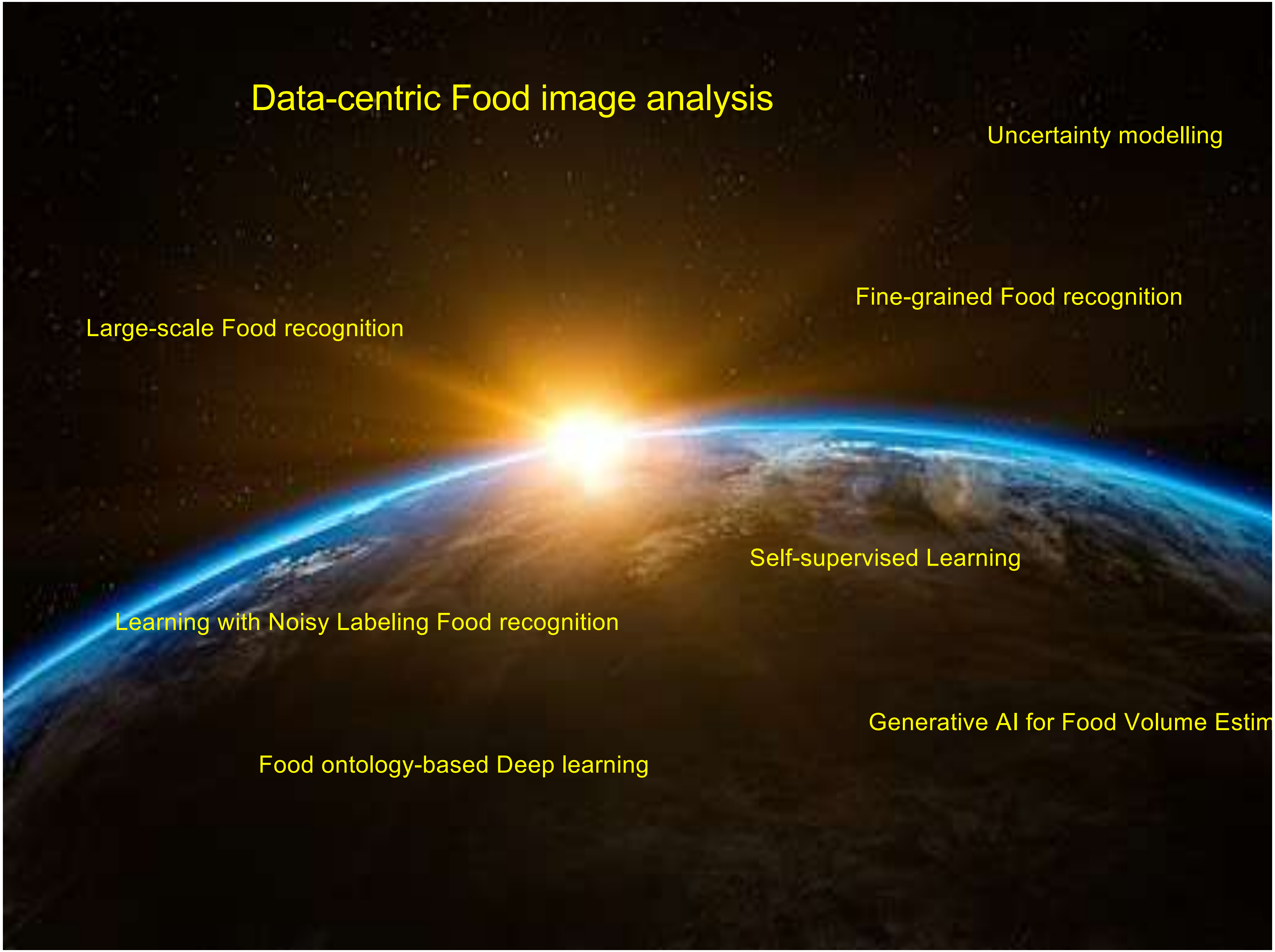
Food1K
Classes: 1000
370K images

Food DB	ImageNet	Future Food DB
150.000 images 231 categories	1.400.000 images 1000 categories	????? images 200.000 categories

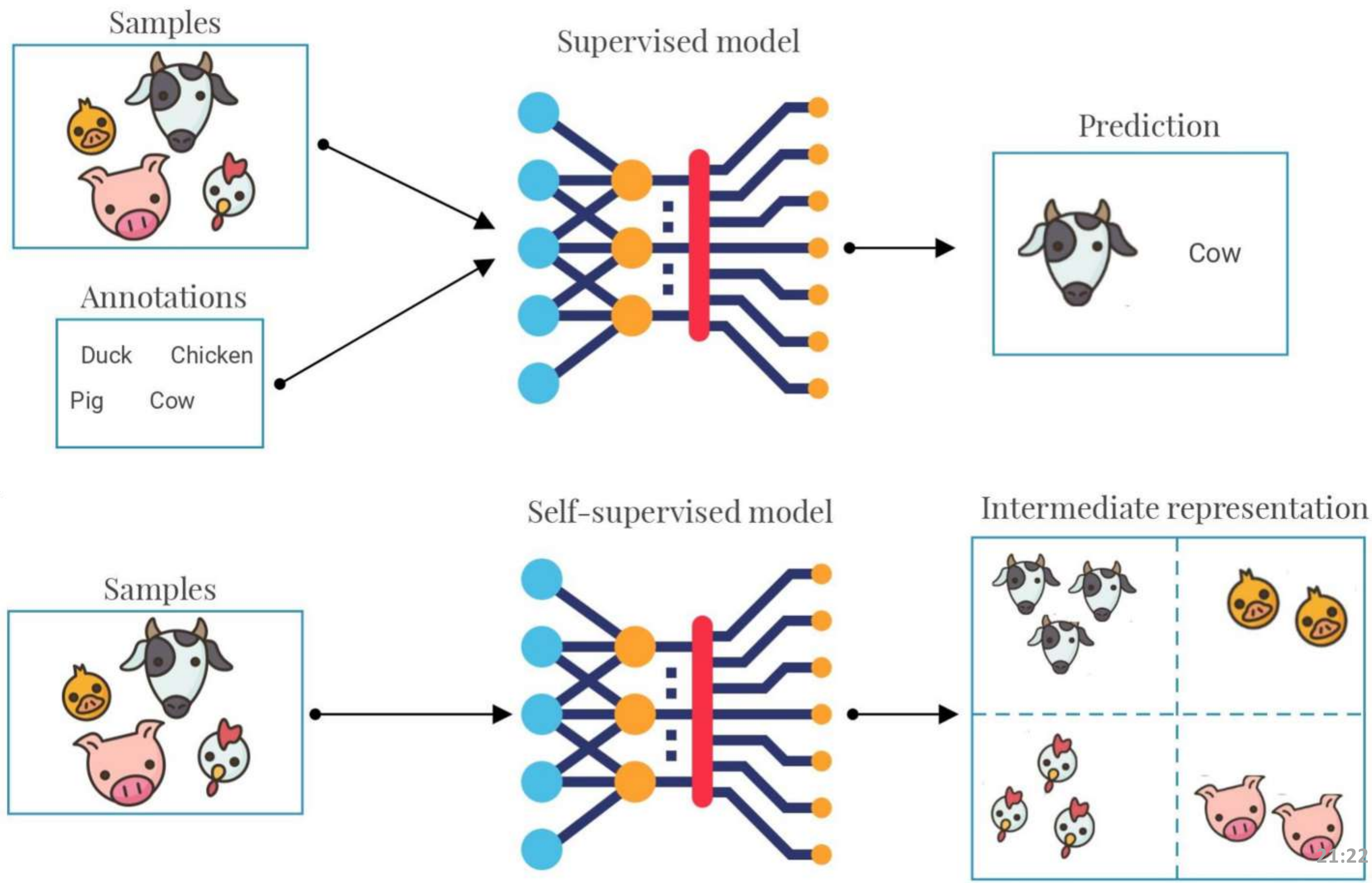
Current SoA on Food recognition

- 79% on UECFOOD
- 44% on ChinaFood1000

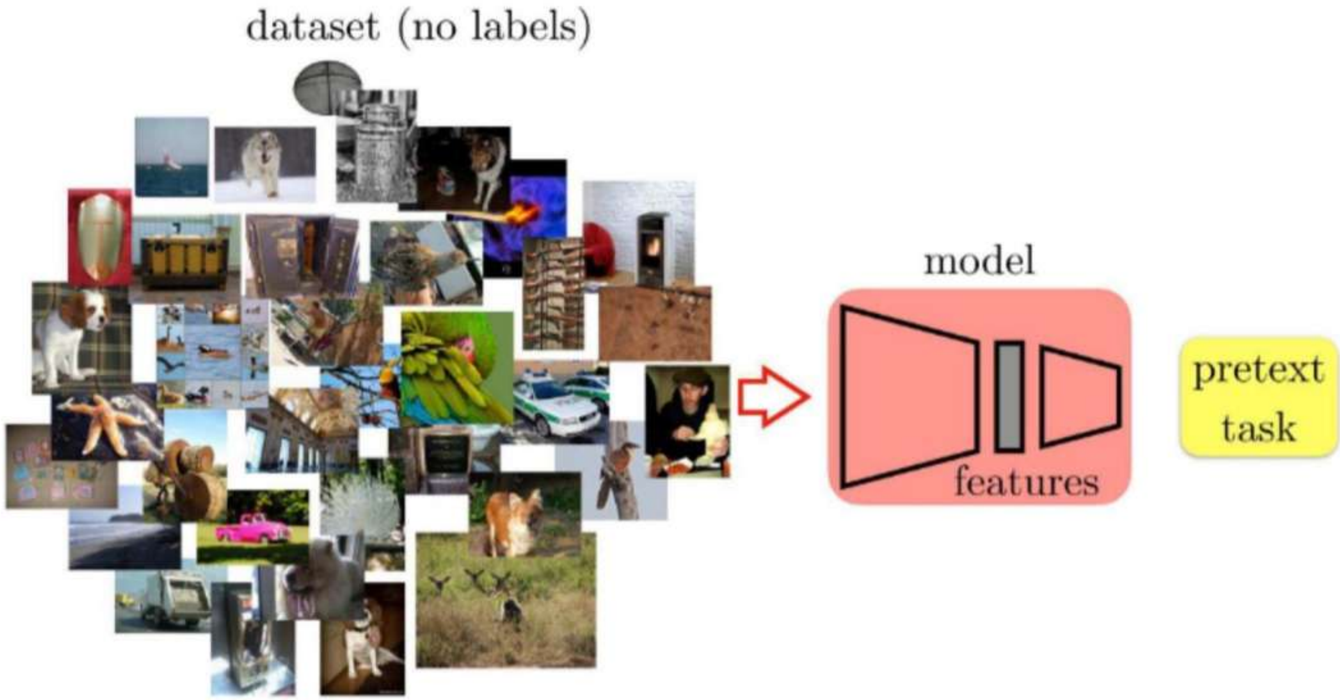
How to leverage from the huge amount of non-annotated data/images?


Data-centric Food image analysis
Uncertainty modelling
Fine-grained Food recognition
Large-scale Food recognition
Self-supervised Learning
Learning with Noisy Labeling Food recognition
Generative AI for Food Volume Estim
Food ontology-based Deep learning

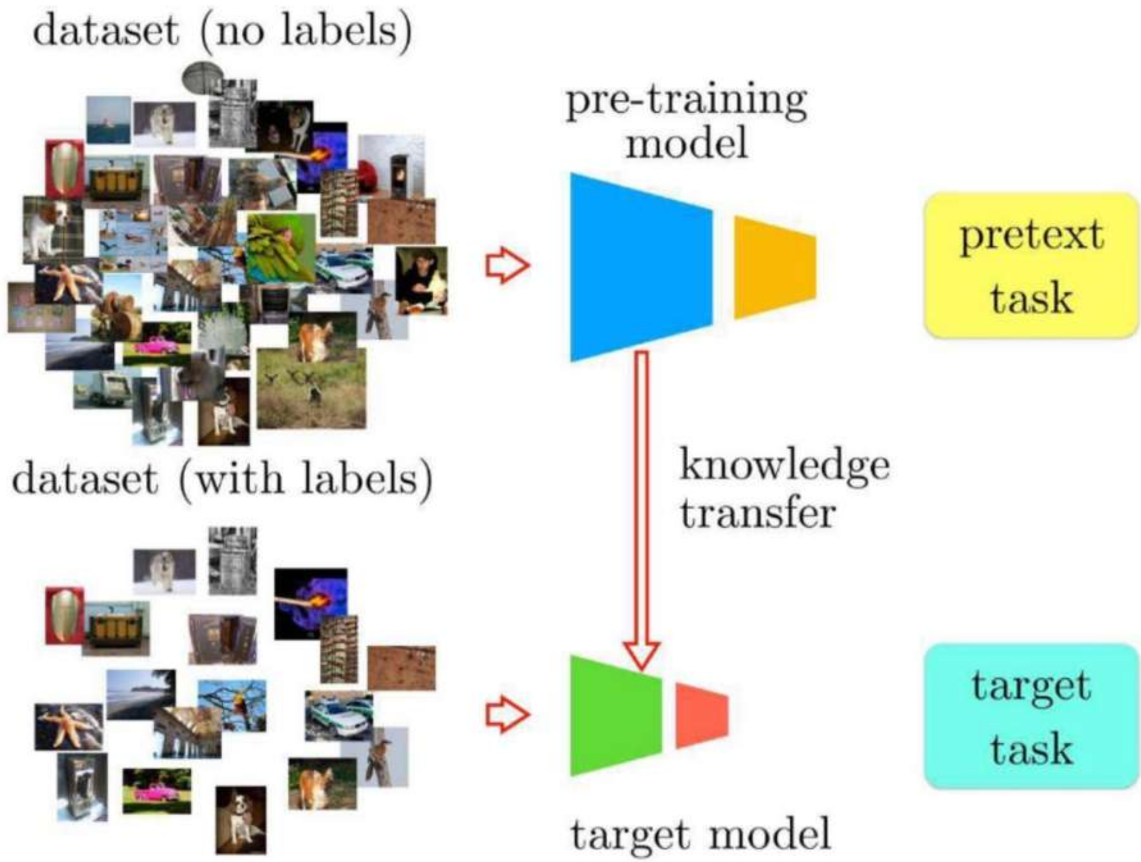
Supervised vs unsupervised learning



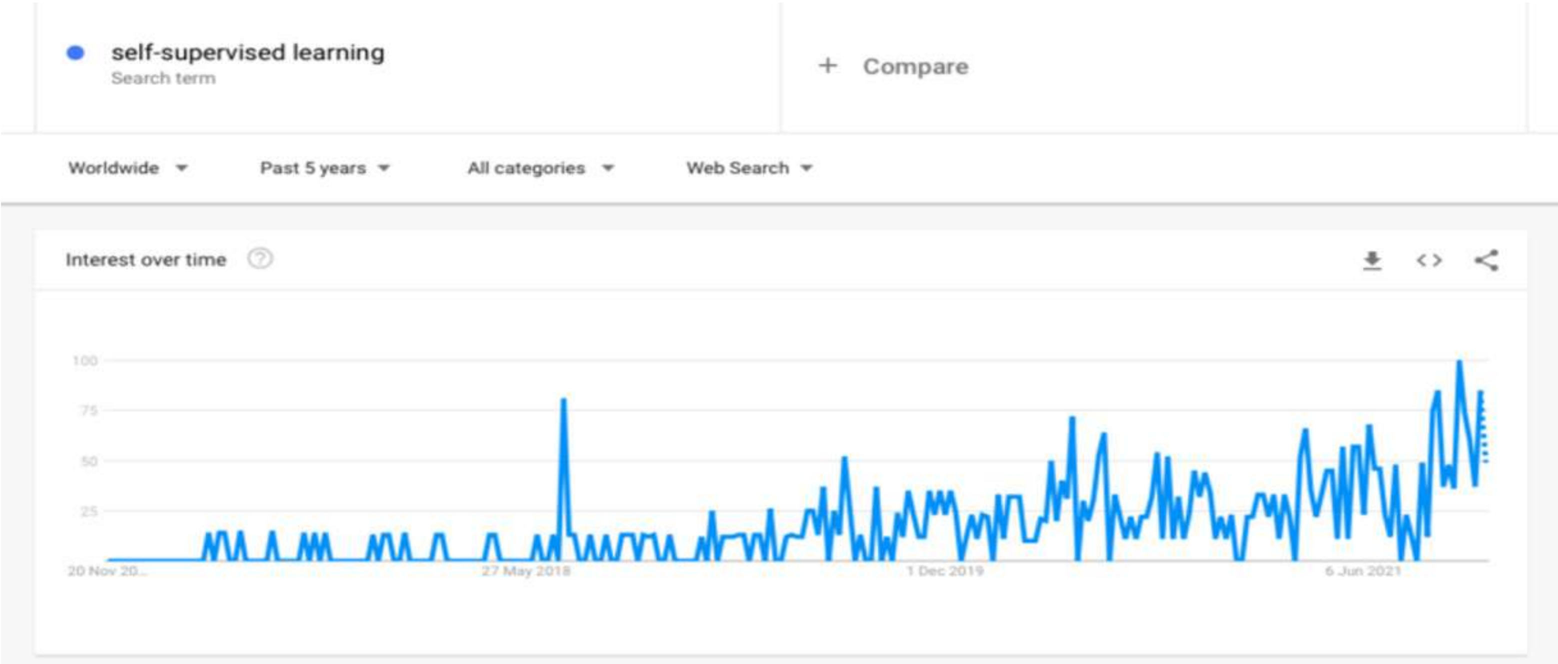
What is self-supervised learning?



What is self-supervised learning (SSL)?



SSL: Benefits & Uses in 2023



https://research.aimultiple.com/self-supervised-learning/

Self-Taught AI Shows Similarities to How the Brain Works

SSL allows a neural network to **figure out for itself what matters**.

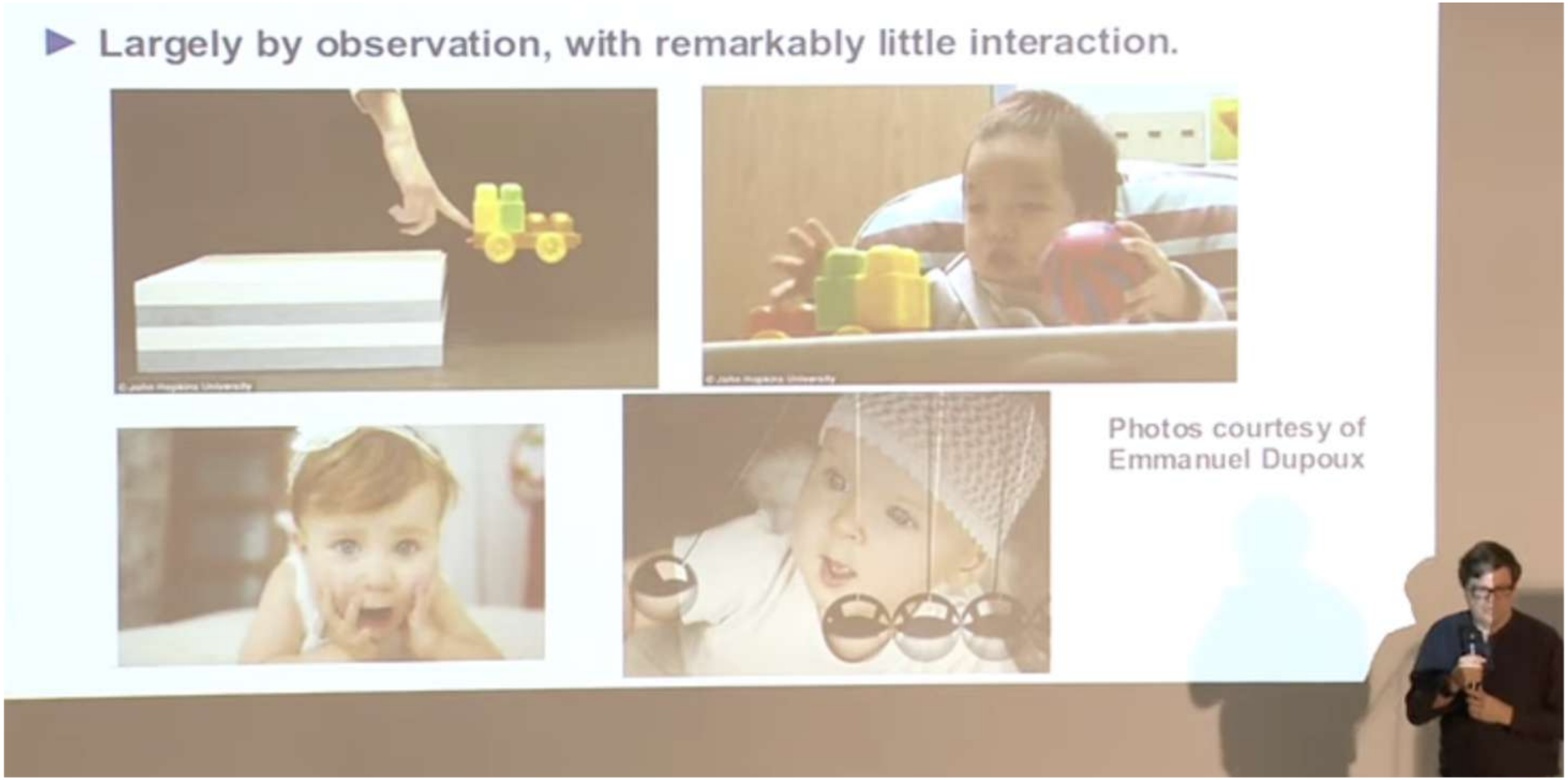
Explore neural networks trained **with little or no human-labelled data.**

Computational **models of the mammalian visual and auditory systems** built using self-supervised learning models have shown a **closer correspondence to brain** function than their supervised-learning counterparts.

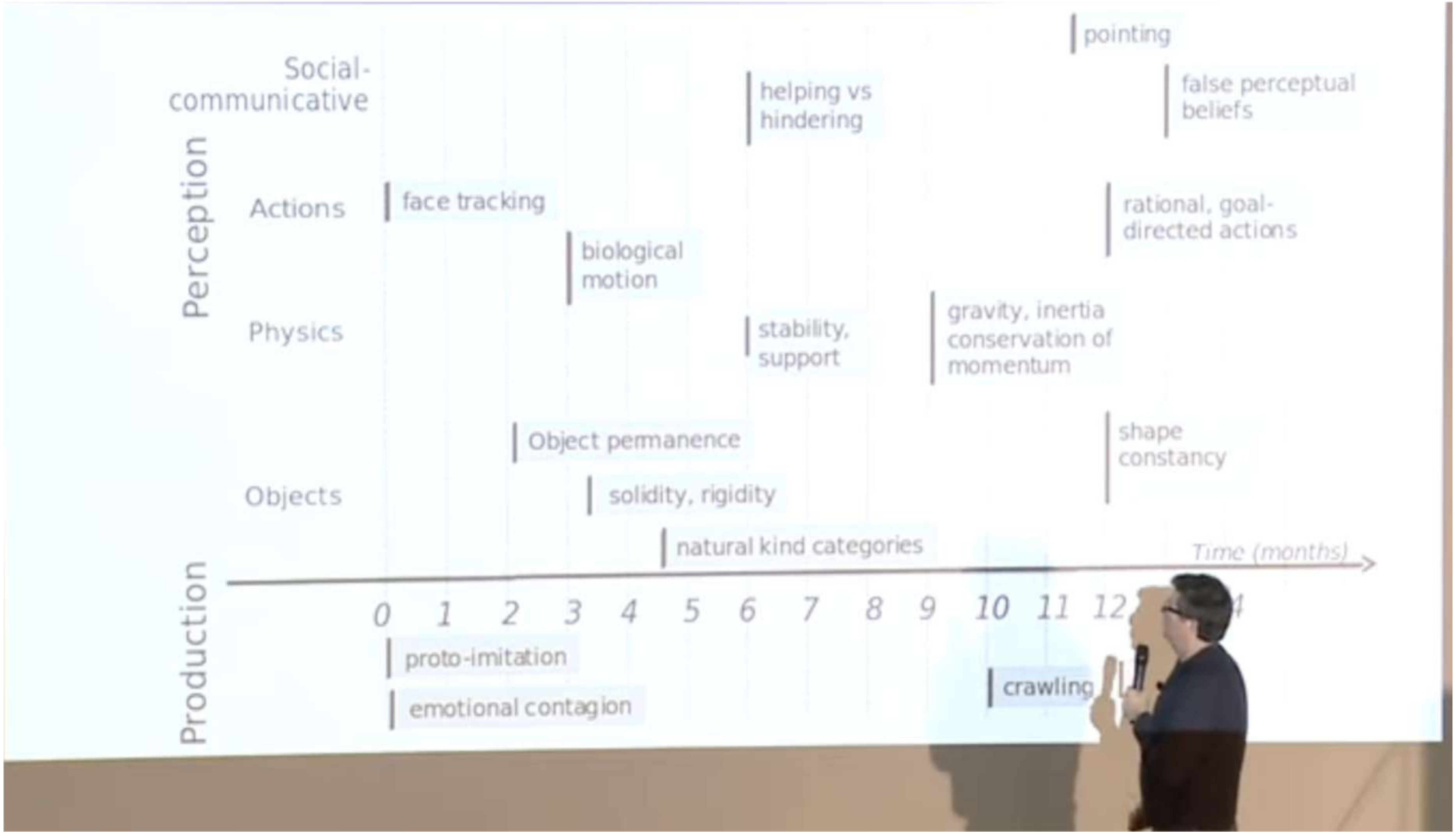
Alexei Efros, University of California, Berkeley, “Most modern AI systems are too reliant on human-created labels. They don’t really learn the material”.

https://www.quantamagazine.org/self-taught-ai-shows-similarities-to-how-the-brain-works-20220811/

Babies learn how the world works by observation



Early conceptual acquisition in infants (from Emmanuel Dupoux)



Artifical vs Natural NNs

Understand brain through NNs:

- the brain is full of feedback connections, while current models have few such connections, if any.

An obvious next step: use SSL to train highly recurrent networks and see how the activity in NNs compares to real brain activity.

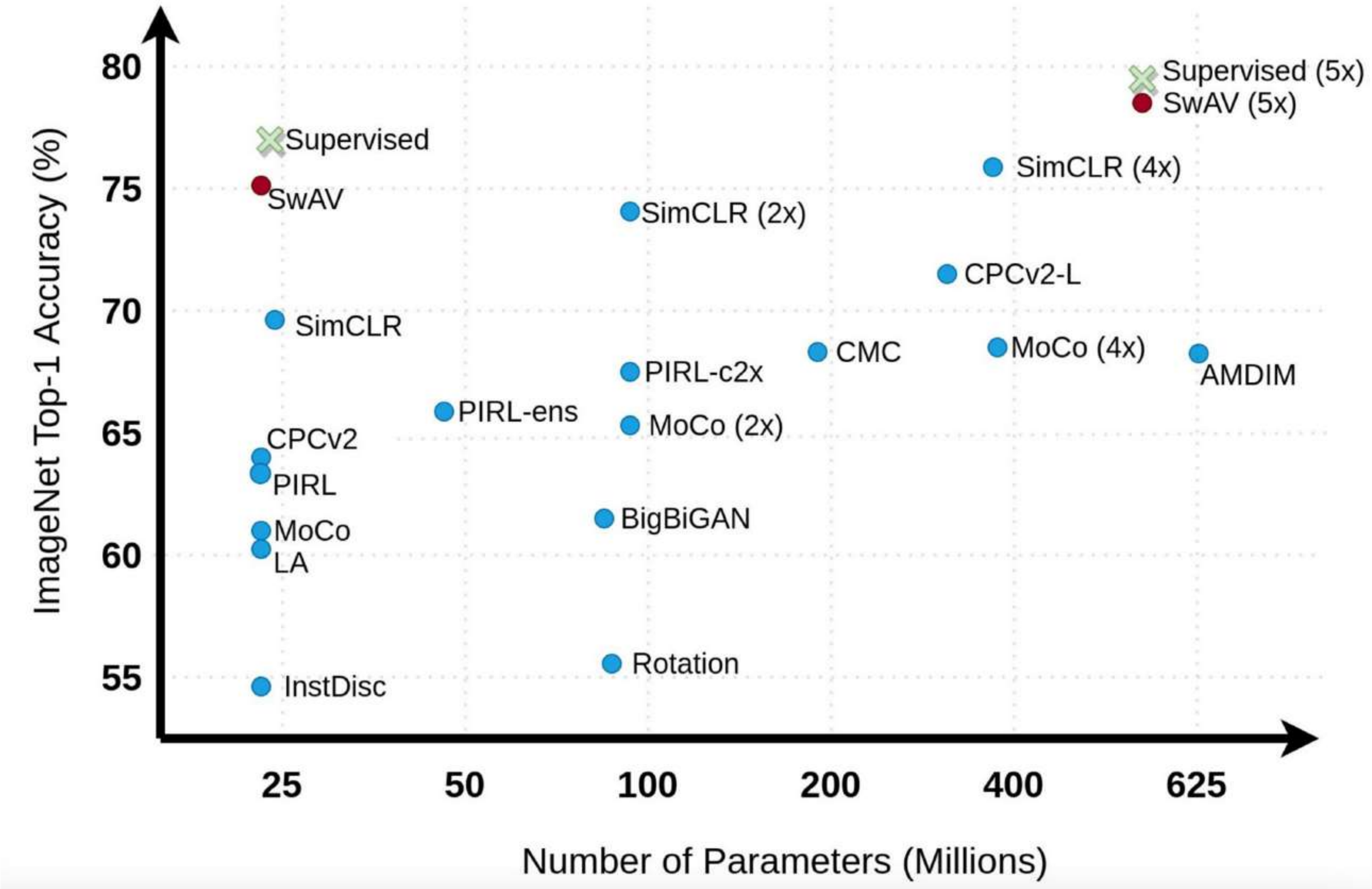
Crucial step: match the activity of NNs in SSL models to the activity of individual biological neurons.

“No doubt that 90% of what the brain does is self-supervised learning,” Blake Richards, a computational neuroscientist at McGill University and Mila, the Quebec Artificial Intelligence Institute.

Hypothesis: the visual systems of humans and other primates are the best studied of all animal sensory systems,

- but neuroscientists have struggled to explain why they include two separate pathways:
  - the ventral visual stream, which is responsible for recognizing objects and faces, and
  - the dorsal visual stream, which processes movement (the “what” and “where” pathways, respectively).

Self-supervised vs supervised learning

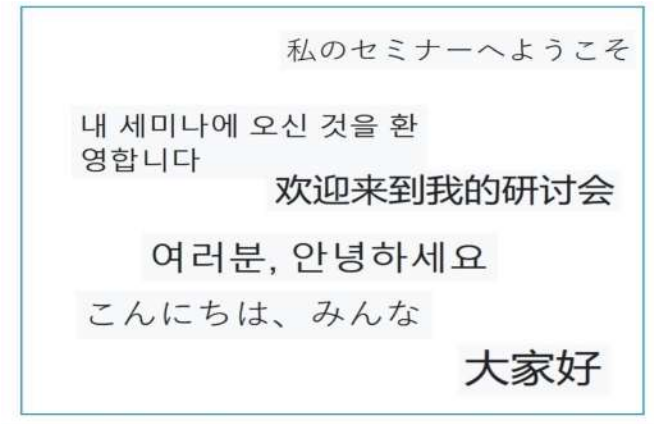


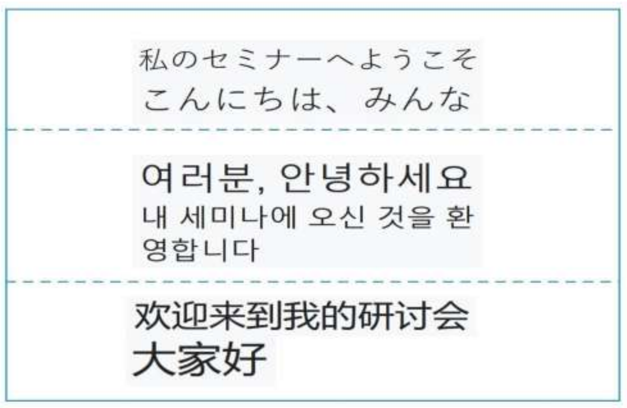
JAISWAL, Ashish, et al. A survey on contrastive self-supervised learning.Technologies, 2020, 9.1: 2.

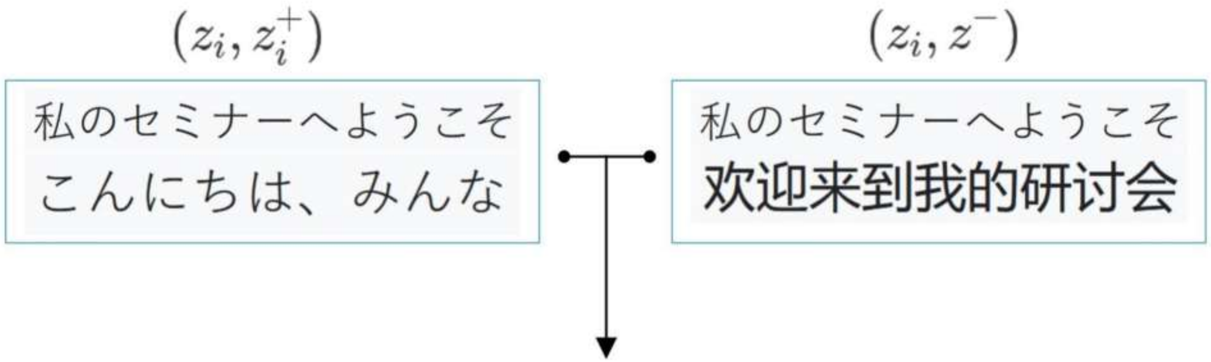
What can you do if you have a lot of just data?

What can you do if you have a lot of just data and may be a not-trained model?

What can you do if you have a lot of just data and may be a not-trained model?



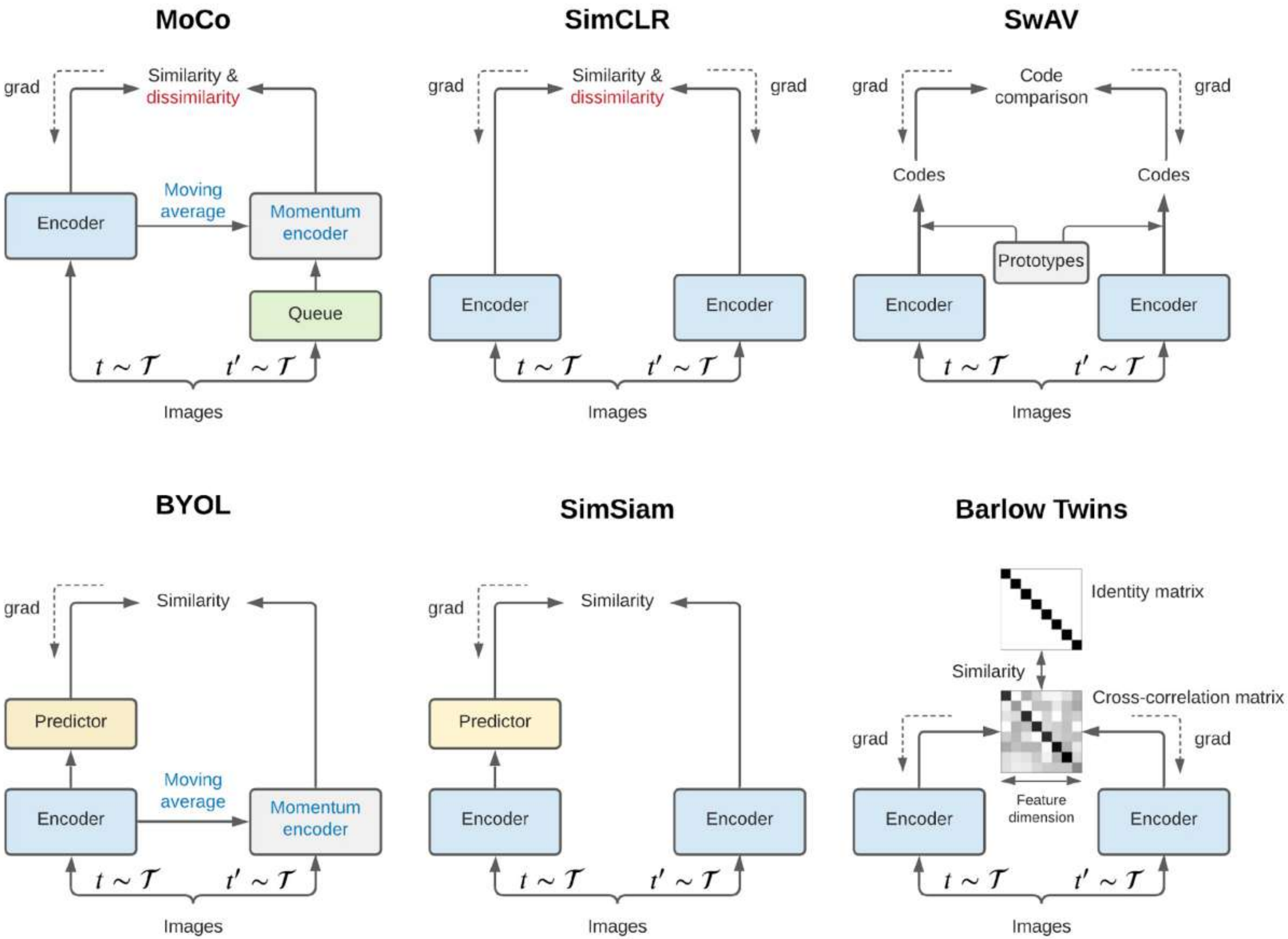




Contrastive loss

$$\mathcal{L}_i^{\text{InfoNCE}} = -\log \frac{\exp\left(z_i \cdot z_i^+/\tau\right)}{\exp\left(z_i \cdot z_i^+/\tau\right) + \sum_{z^- \in \mathcal{N}_i} \exp\left(z_i \cdot z^-/\tau\right)}$$

State-of-the-art Contrastive SSL models



Momentum Contrasting (MOCO)

Given an image x_i, MoCo learns a query encoder $q = f_q(x_i)$ able to **differentiate $q_i = f_q(x_i)$ from the other images**.

Positive pairs: 2 representations of the same image without augmentation.

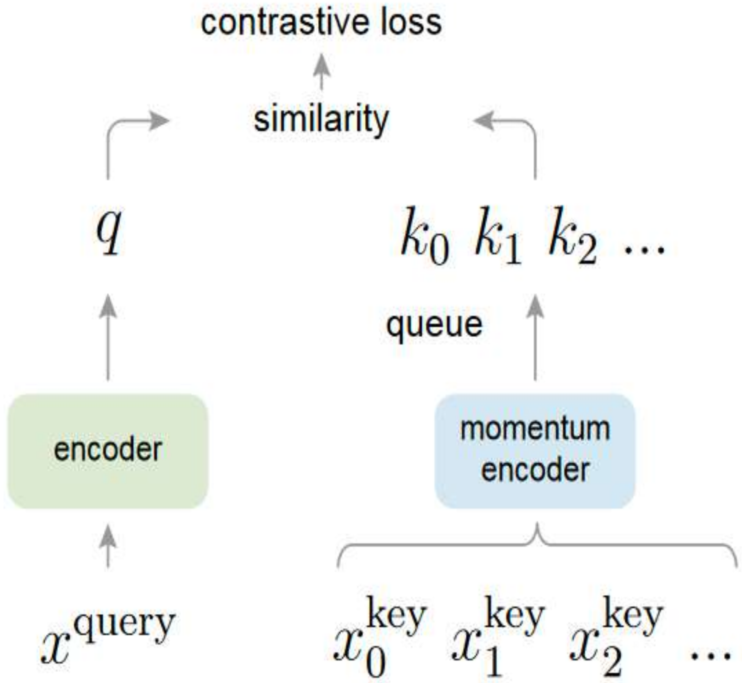
An asynchronously updated momentum encoder $f_k(.)$ is used to generate the positive counterpart $k^+ = f_k(x_i)$.

Negative samples: MoCo derives from a memory bank, storing previously encoded representations.

The model optimizes the following objective function:

$$L_i^{MoCo} = -log(\frac{exp(q_i \cdot k_i^+/\tau)}{\sum_{k=1}^{K} exp(q_i \cdot k_k^-/\tau)})$$

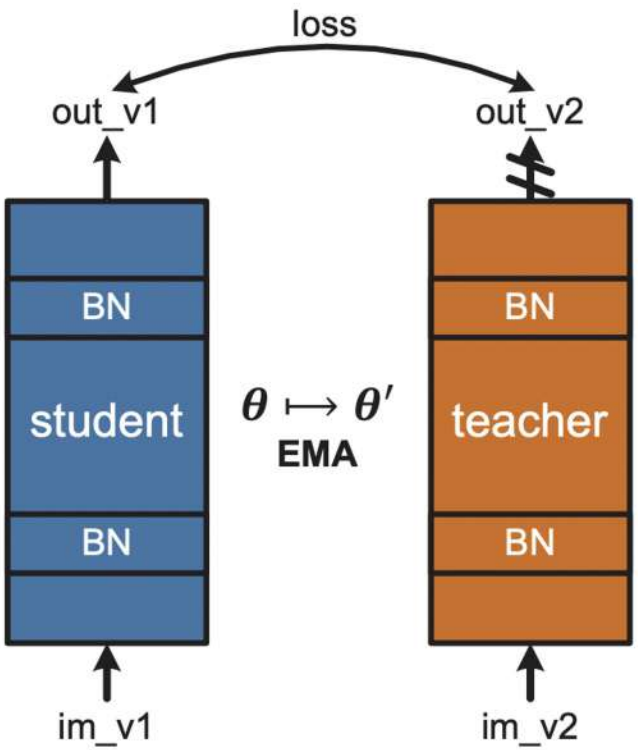
where K is the number of negative samples in the queue.



Kaiming He et al. "Momentum Contrast for Unsupervised Visual Representation Learning", CVPR 2020, pp. 9729–9738.

Exponential Moving Average Prediction

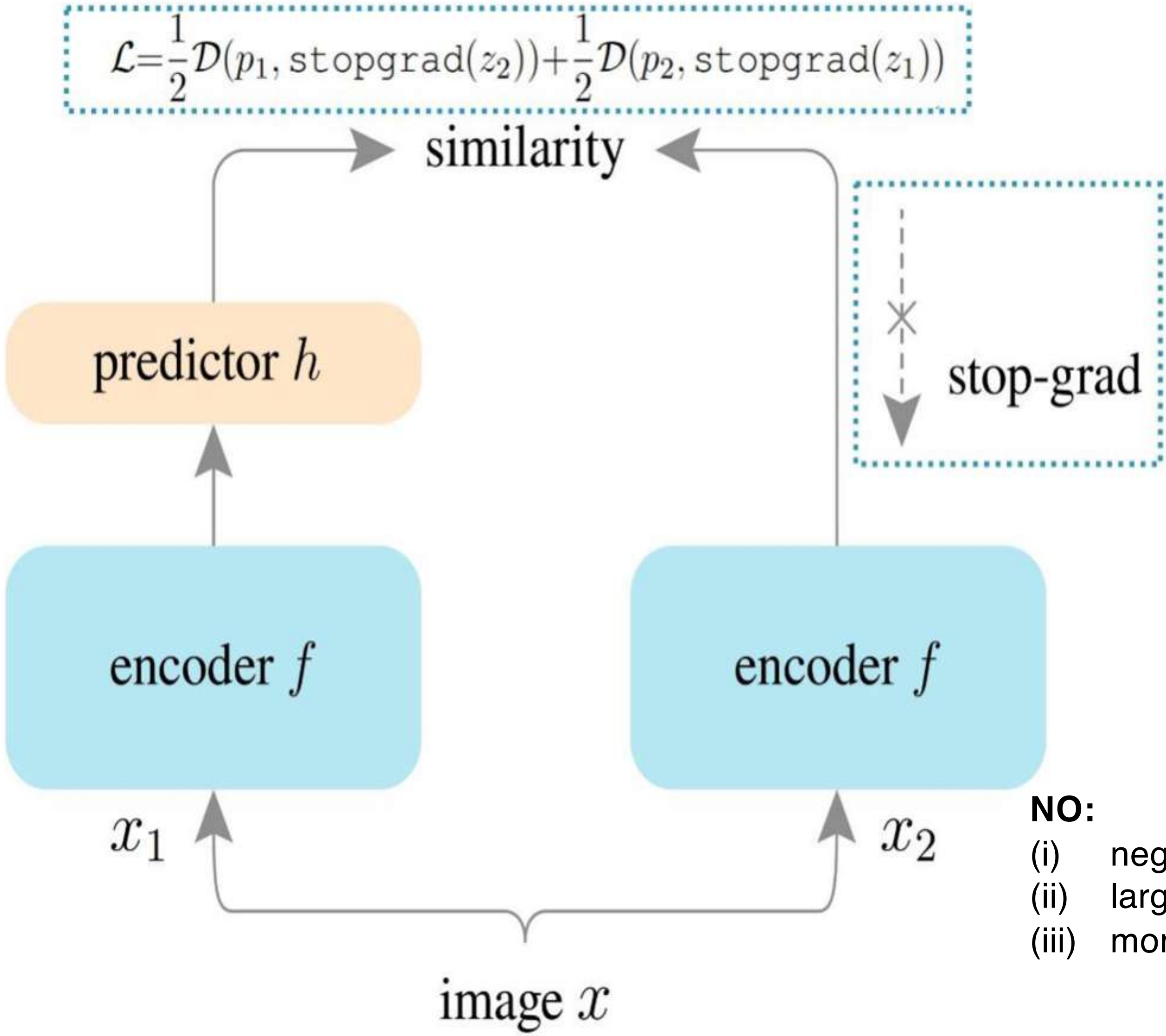
EMA denotes exponential moving average updates.



$$\xi \leftarrow \tau\xi + (1 - \tau)\theta$$

The EMA-teacher framework with standard Batch Normalization.

SSL Framework: SimSiam

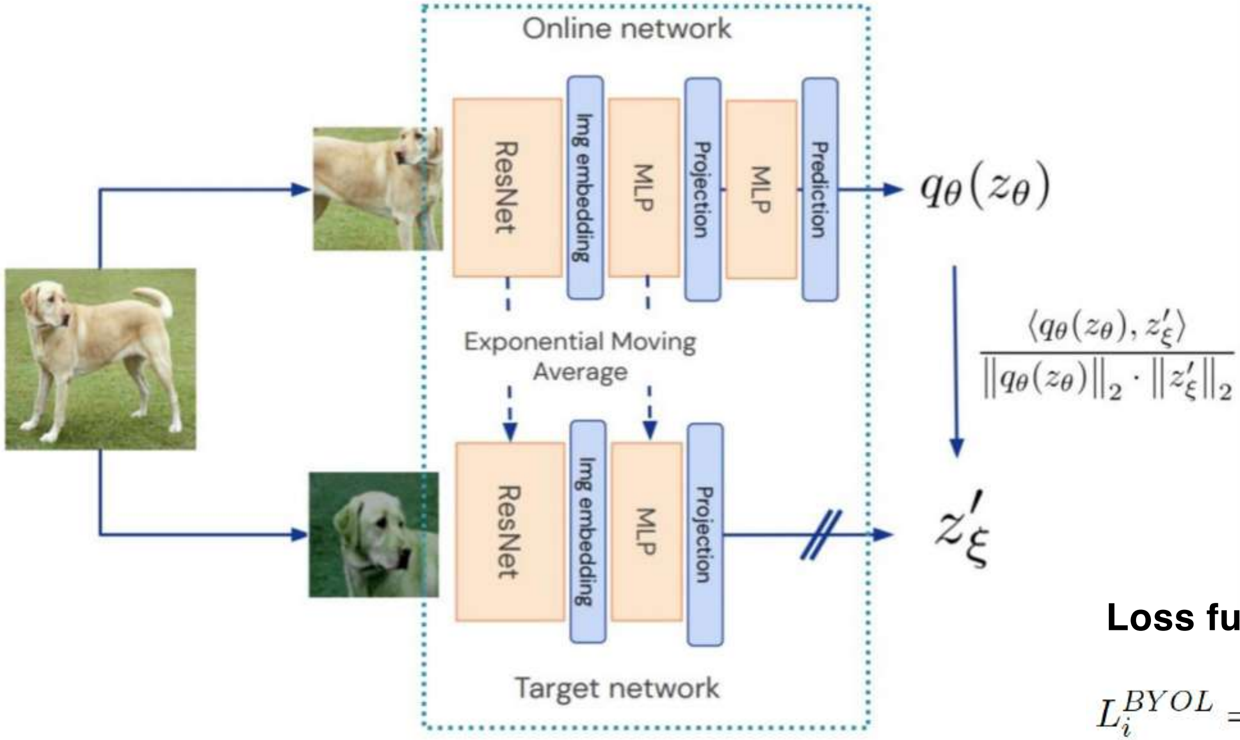


NO:

(i) negative sample pairs,
(ii) large batches,
(iii) momentum encoders.

X. Chen and K. He. "Exploring Simple Siamese Representation Learning". CVPR, 2021, pp. 15750–15758

SSL Framework: BYOL



Avoid negative sampling!

Introduces predictor:

It tries to predict different views (regression targets) of the same image directly in the representation space using a predictor.

Loss function:

$$L_i^{BYOL} = \|p_i^1 - z_i^2\|_2^2 = 2 - 2 \cdot \frac{\langle p_i^1, z_i^2 \rangle}{\|p_i^1\|_2 \cdot \|z_i^2\|_2}.$$

J.-B. Grill et al. "Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning". NIPS. Vol. 33. 2020, pp. 21271–21284

SimCLR by the Google AI team

Introduces projectors: a learnable nonlinear transformation between the representation and the contrastive loss

Positive sampling: Given a batch of N samples, the pretext task P generates two augmented views x^a_i and x^+_i for each sample x_i of the batch.

Negative sampling: the rest of the images x^-_i on the same batch to form the negative pairs (x^a_i, x^-_i).

Batch sizes of 8196 are used.

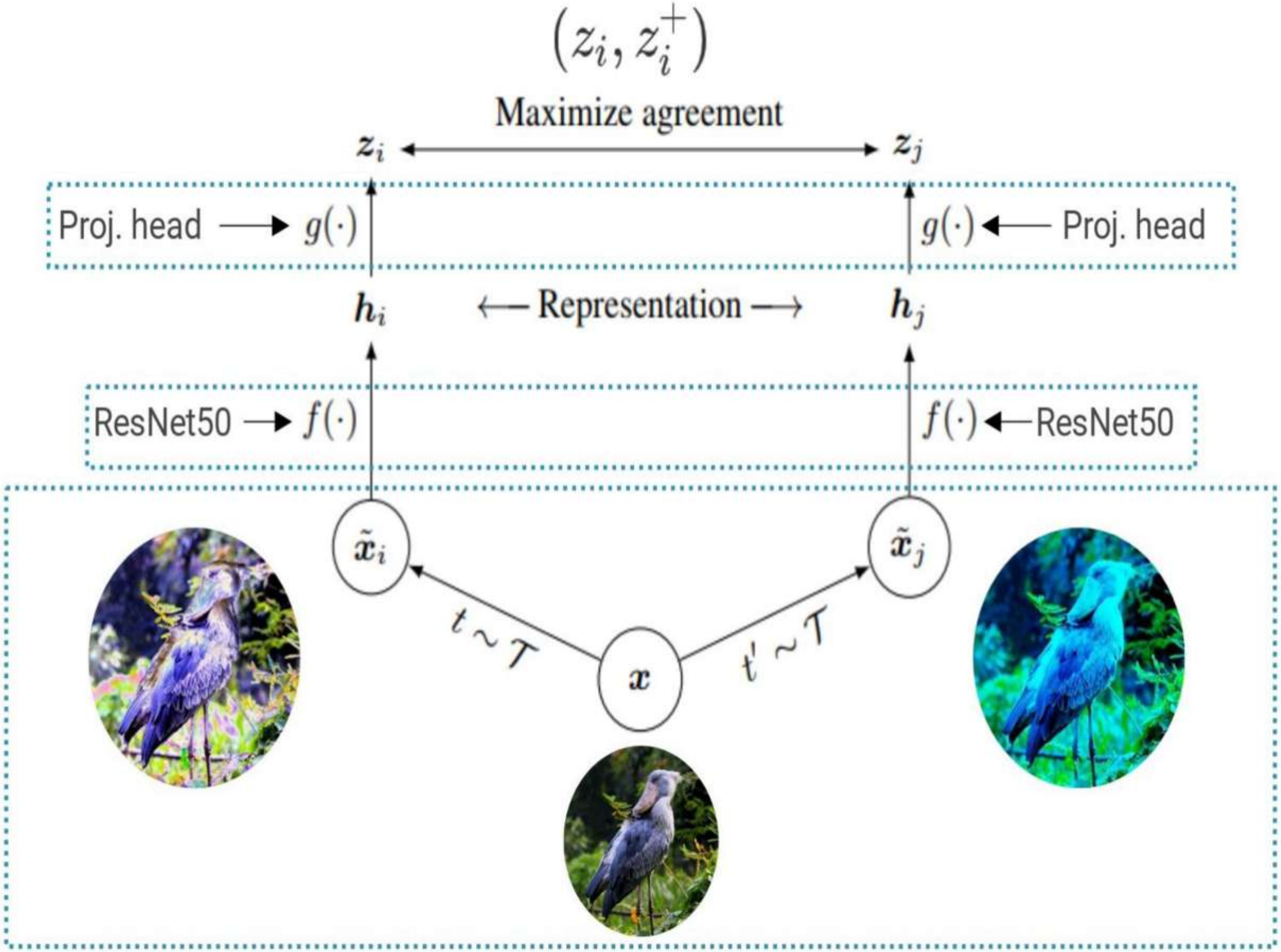
Loss function:

$$L_i^{SimCLR} = -log(\frac{exp(z_i^a \cdot z_i^+ / \tau)}{\sum_{k=1}^{N} exp(z_i^a \cdot z_k^- / \tau)})$$

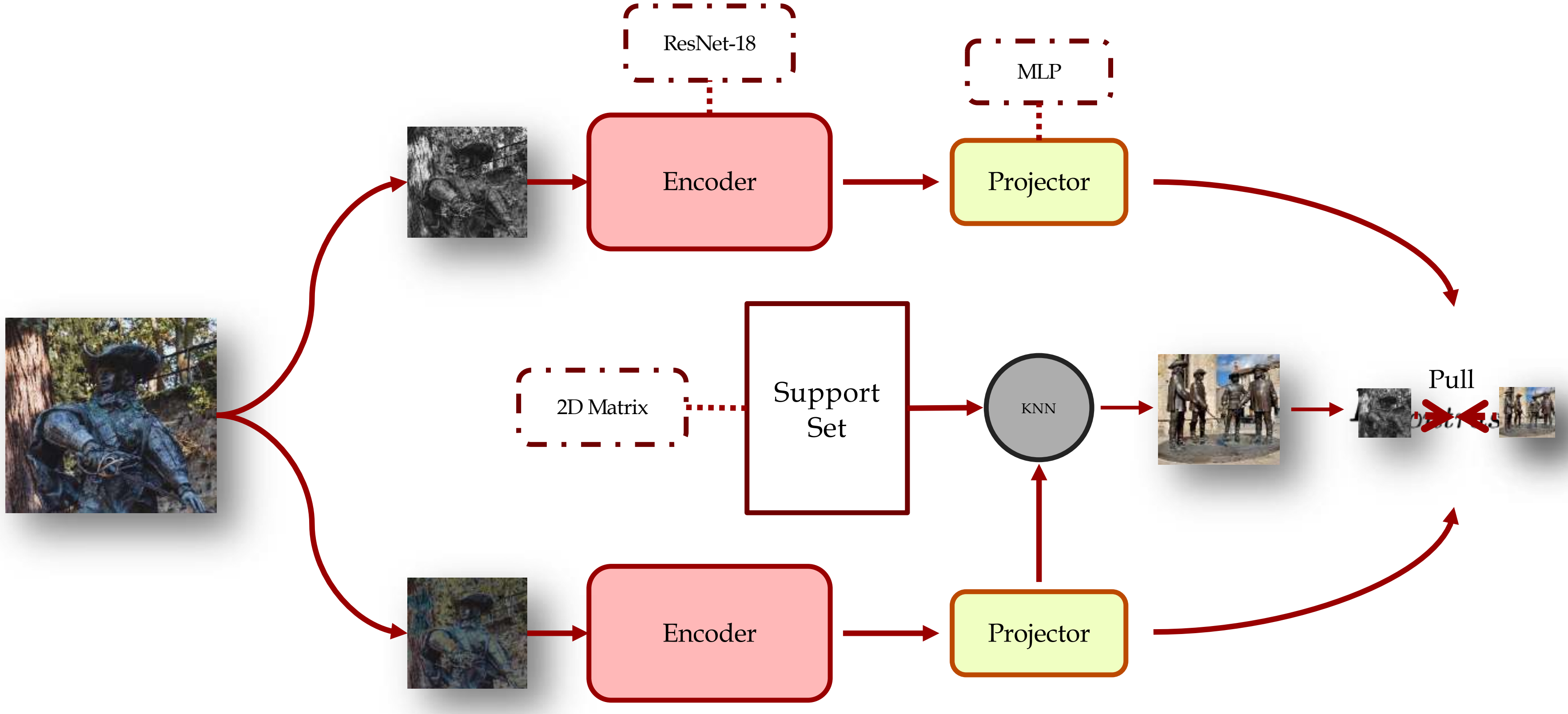
Ting Chen et al. “A Simple Framework for Contrastive Learning of Visual Representations”. 37th ICMLL 2020, pp. 1597–1607.

https://analyticsindiamag.com/what-is-contrastive-self-supervised-learning/

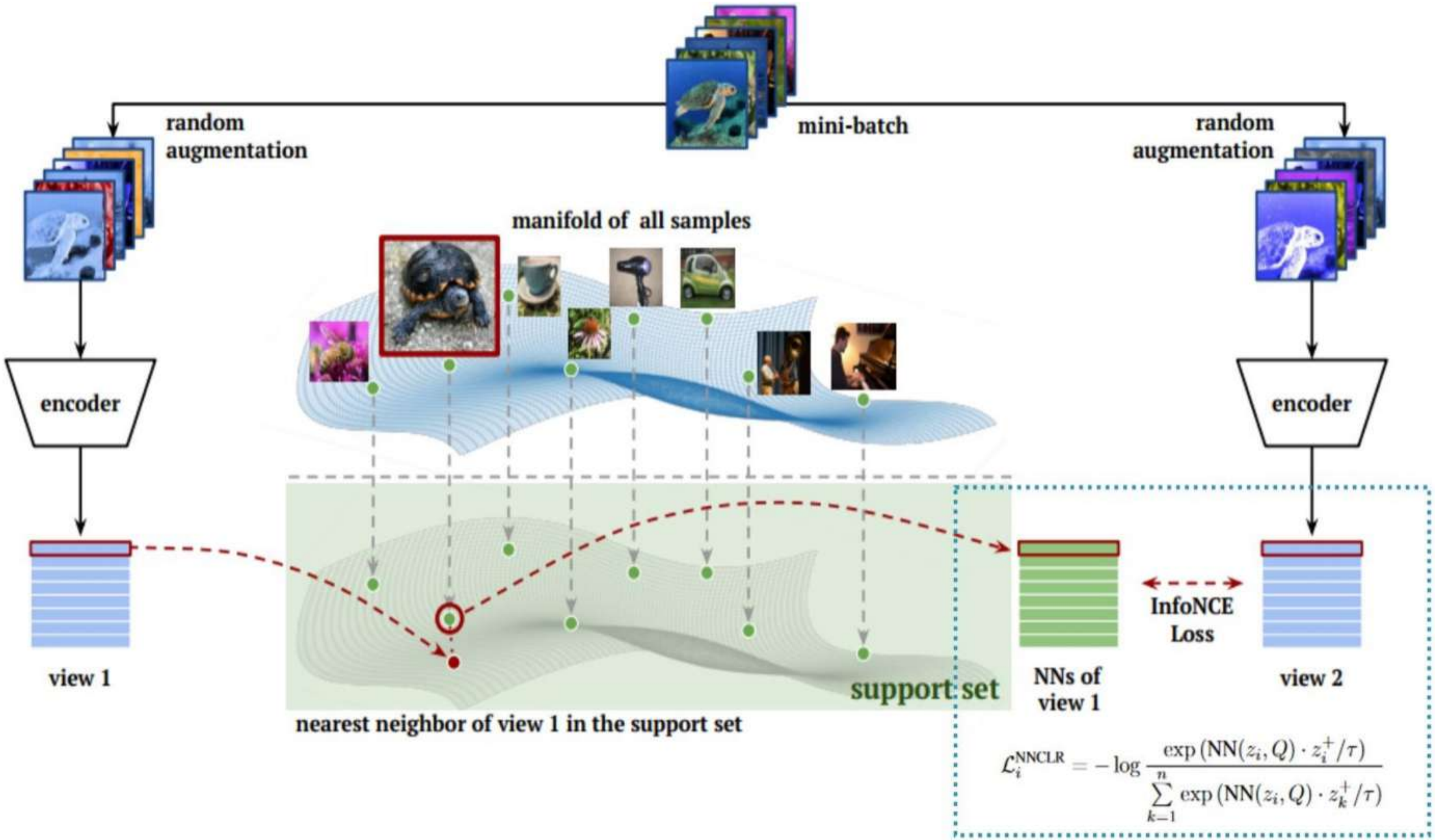
SSL Framework: SimCLR



SimCLR


ResNet-18
MLP
Encoder
Projector
2D Matrix
Support Set
KNN
Pull
Encoder
Projector

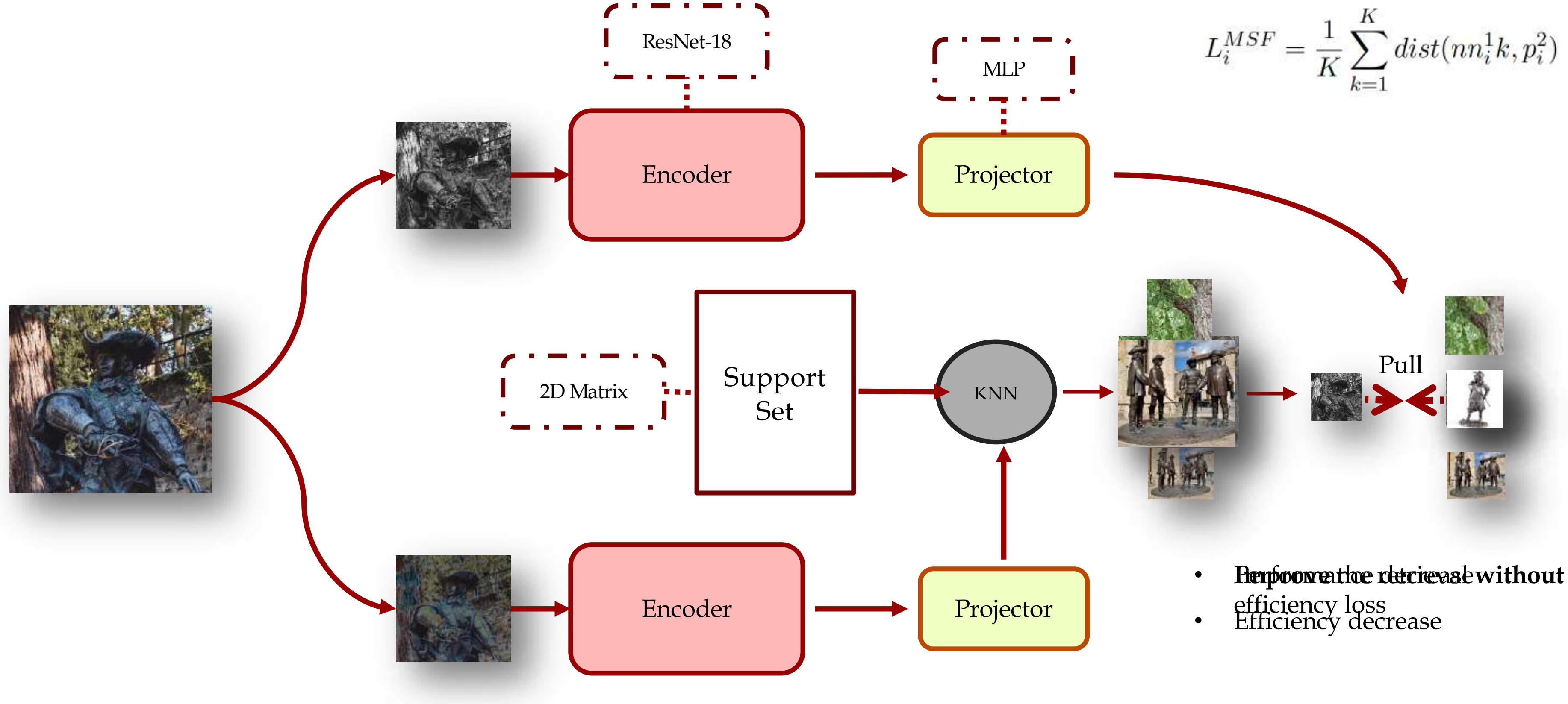
SSL Framework: NNCLR



MSF

SimCLR

Mean shift for Self-supervised learning (MSF)



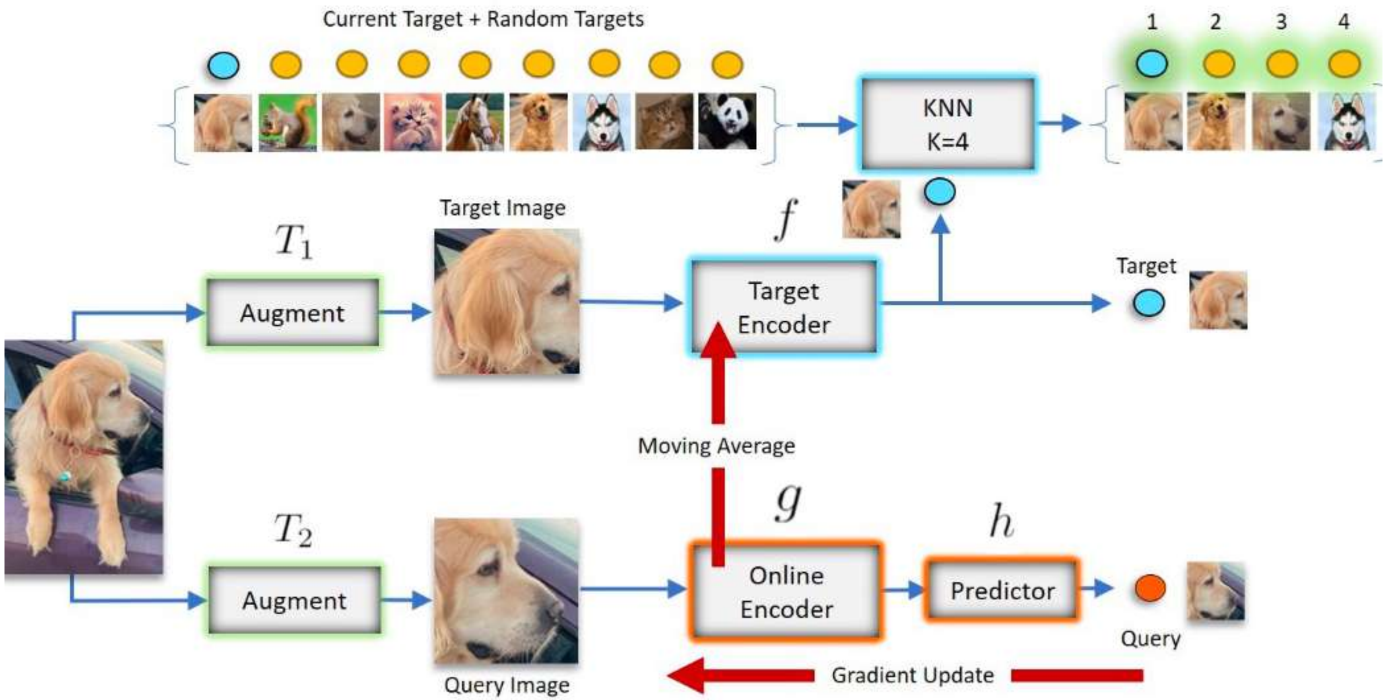
$$L_i^{MSF} = \frac{1}{K}\sum_{k=1}^{K} dist(nn_i^1 k, p_i^2)$$

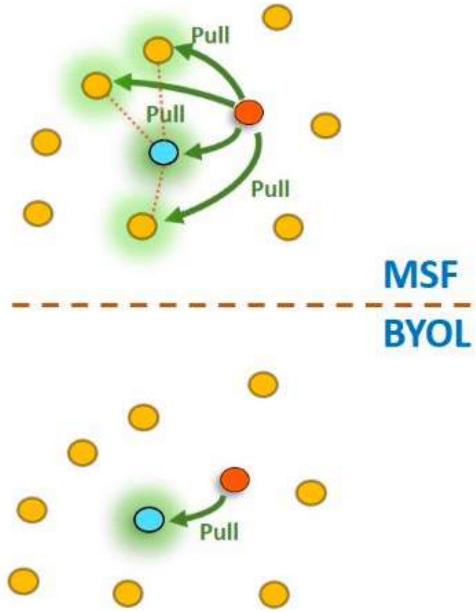
- **Improve the retrieval without efficiency loss**

- Performance decrease
- Efficiency decrease

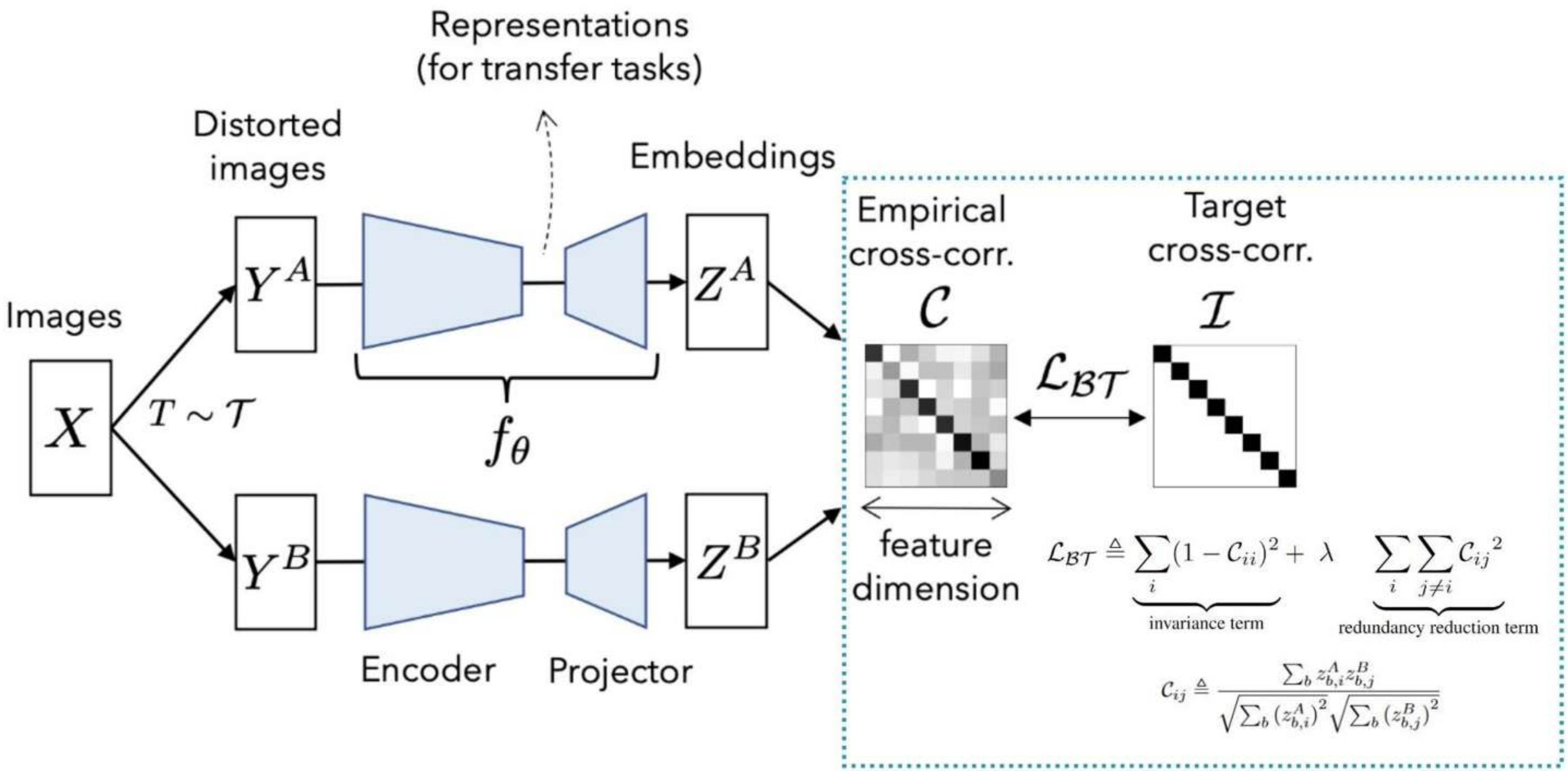
S. A. Koohpayegani, A. Tejankar, and H. Pirsiavash. "Mean Shift for Self-Supervised Learning", CVPR 2021, pp. 10326–10335.

BYOL vs MSF


Current Target + Random Targets
KNN
K=4
1
2
3
4
Target Image
T_1
Augment
f
Target
Encoder
Target
Moving Average
T_2
Augment
g
Online
Encoder
h
Predictor
Query
Query Image
Gradient Update


Pull
Pull
Pull
MSF
BYOL
Pull

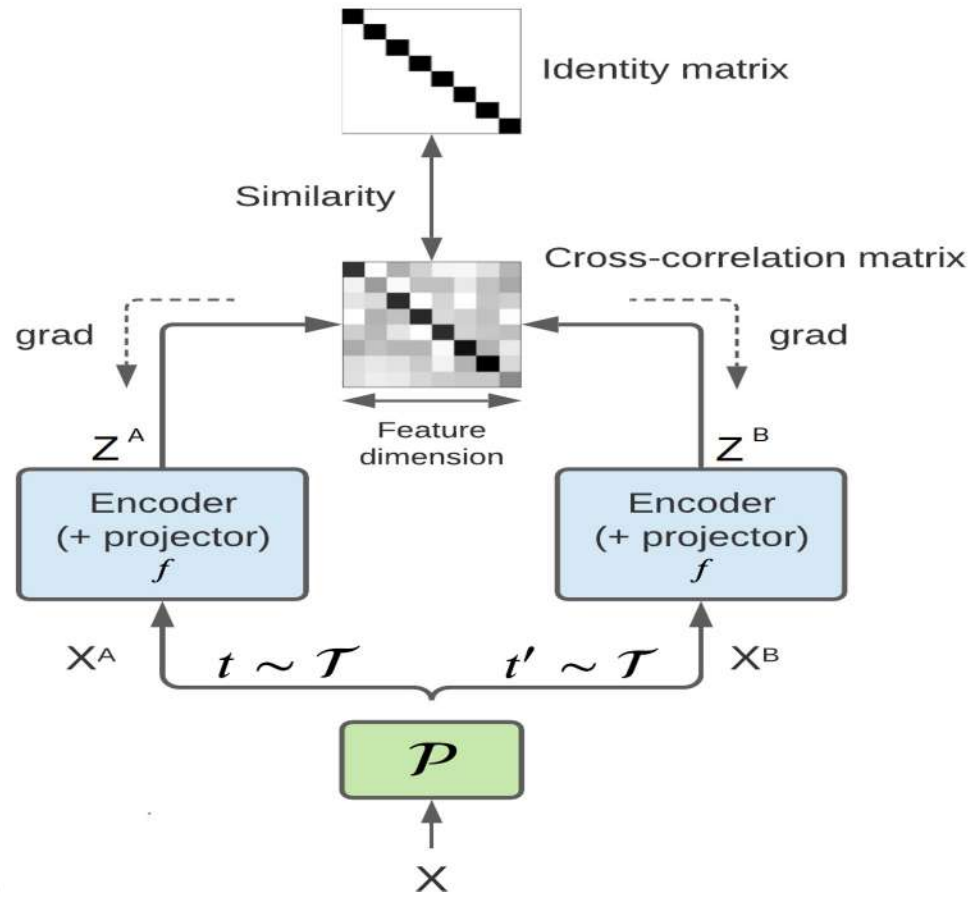
SSL Framework: Barlow Twins



J. Zbontar et al. “Barlow Twins: Self-Supervised Learning via Redundancy Reduction”. 38th ICML. 2021, pp. 12310–12320

Feature Contrasting

Barlow Twins architecture



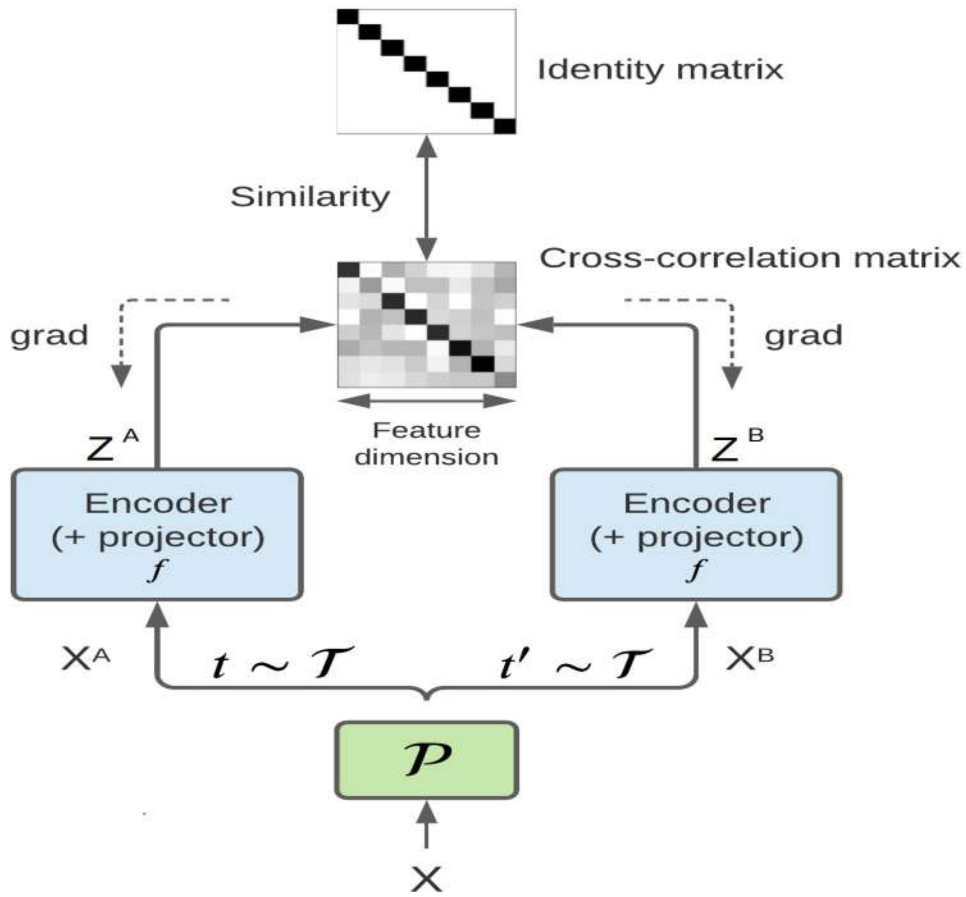
Z^A

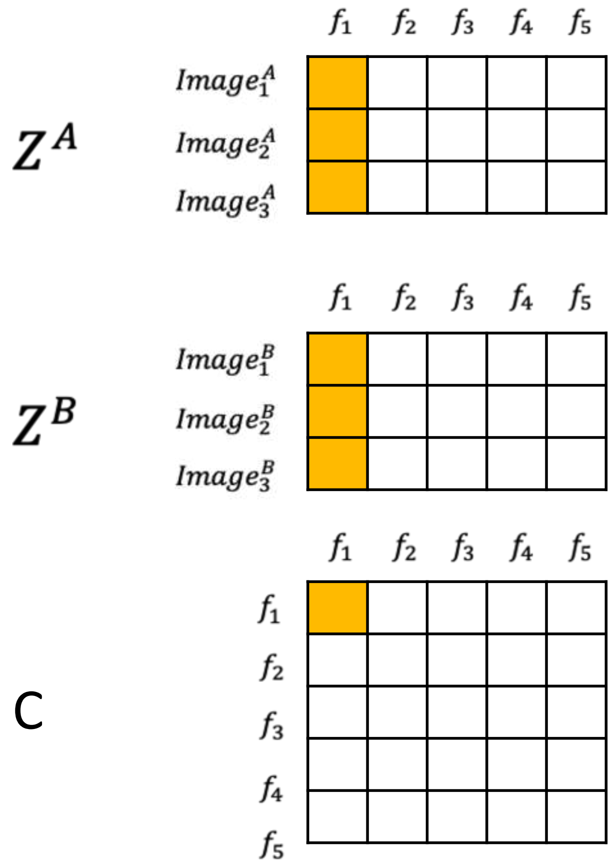
	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

Z^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

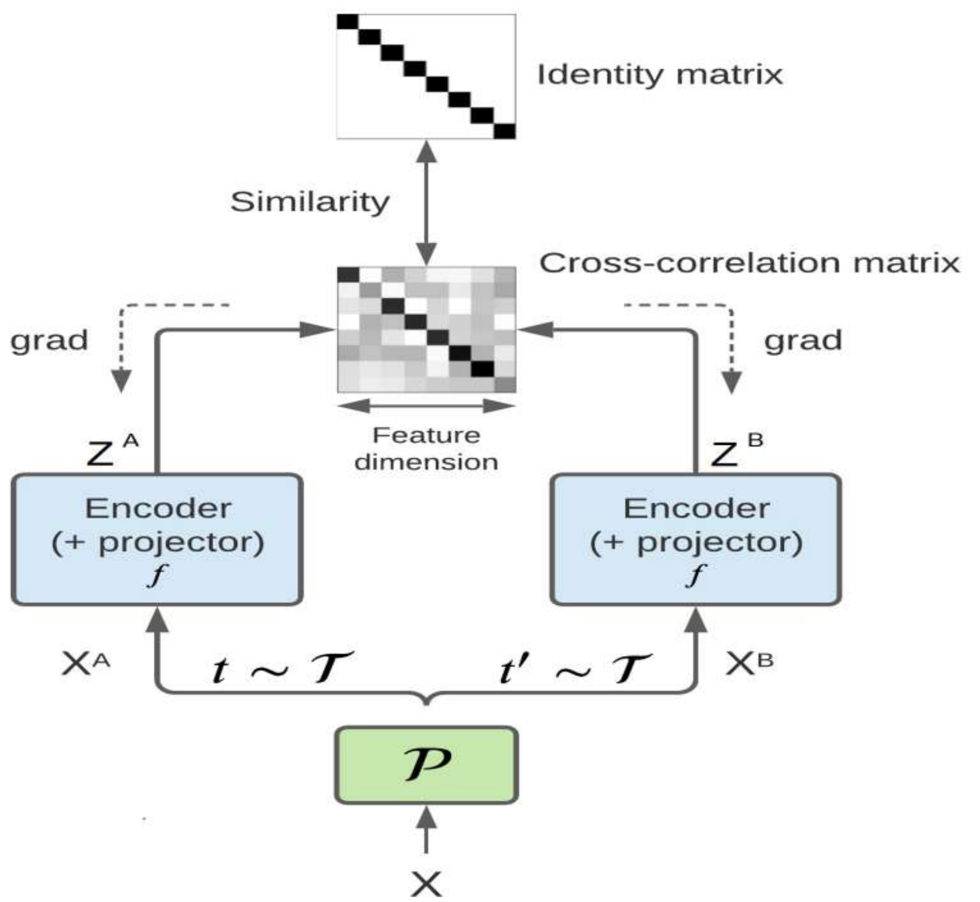
Barlow Twins architecture

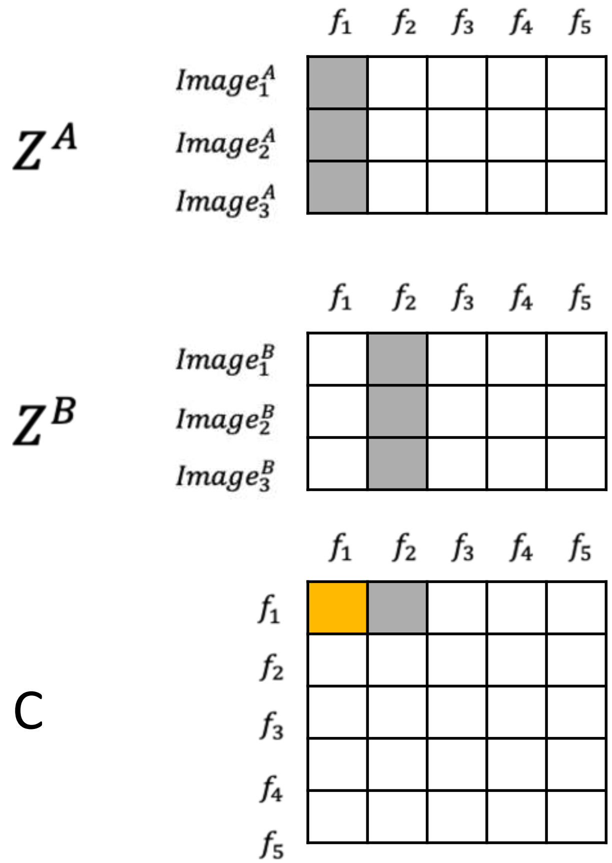




Feature Contrasting

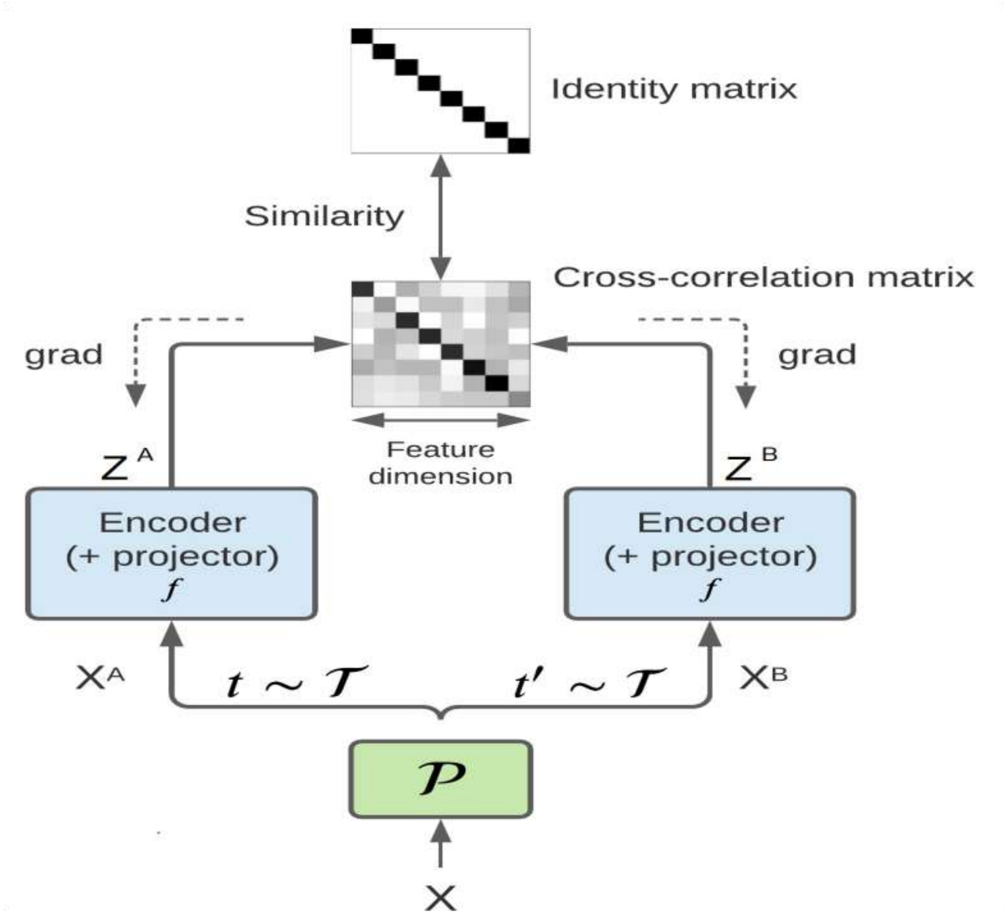
Barlow Twins architecture

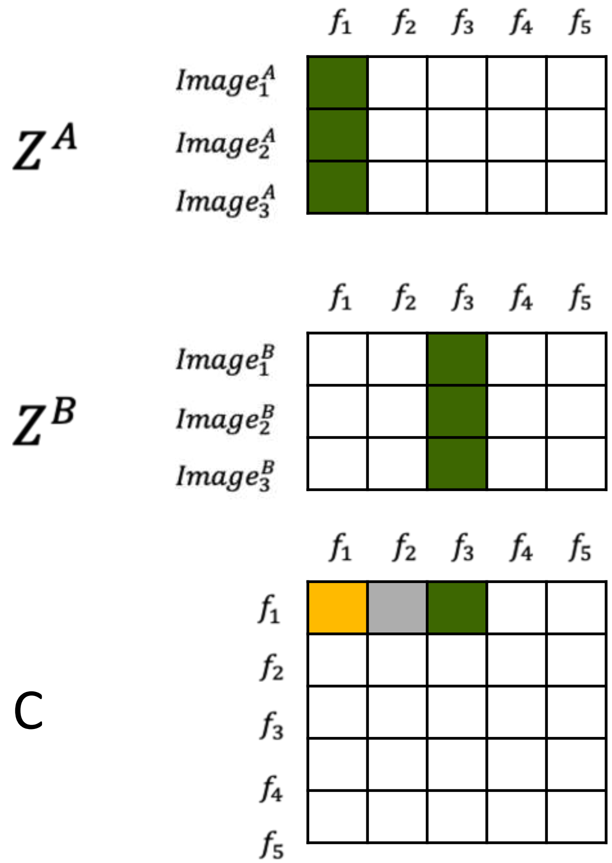




Feature Contrasting

Barlow Twins architecture





Feature Contrasting

C

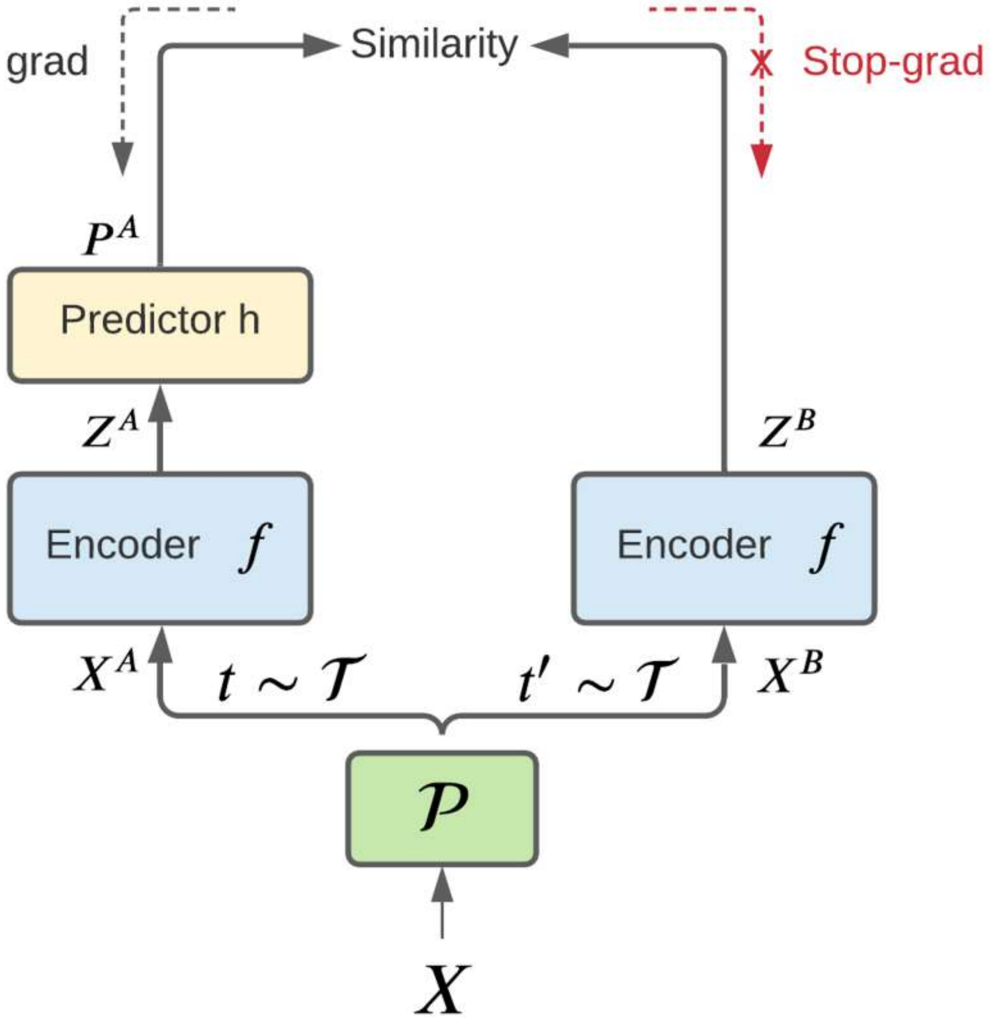
	f_1	f_2	f_3	f_4	f_5
f_1					
f_2					
f_3					
f_4					
f_5					

Barlow Twins' Loss Function

$$\mathcal{L}_{\mathcal{BT}} = \underbrace{\sum_{i=1}^{D} (1 - c_{ii})^2}_{\text{Invariance term}} + \underbrace{\lambda \sum_{i=1}^{D} \sum_{j=1, j \neq i}^{D} c_{ij}^2}_{\text{Redundancy reduction term}}$$

Image Representation Contrasting

Remember: in the SimSiam architecture



P^A

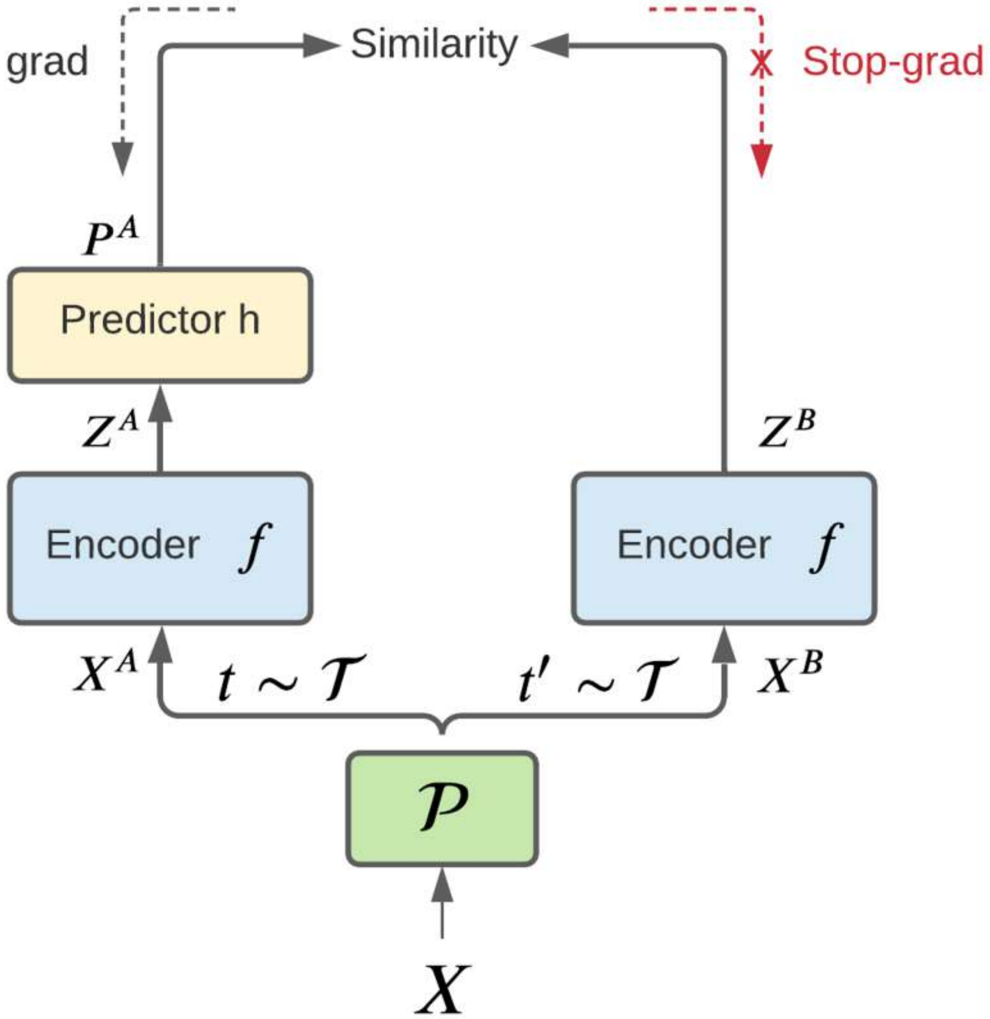
	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

Z^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

Image Representation Contrasting

Remember: in the SimSiam architecture



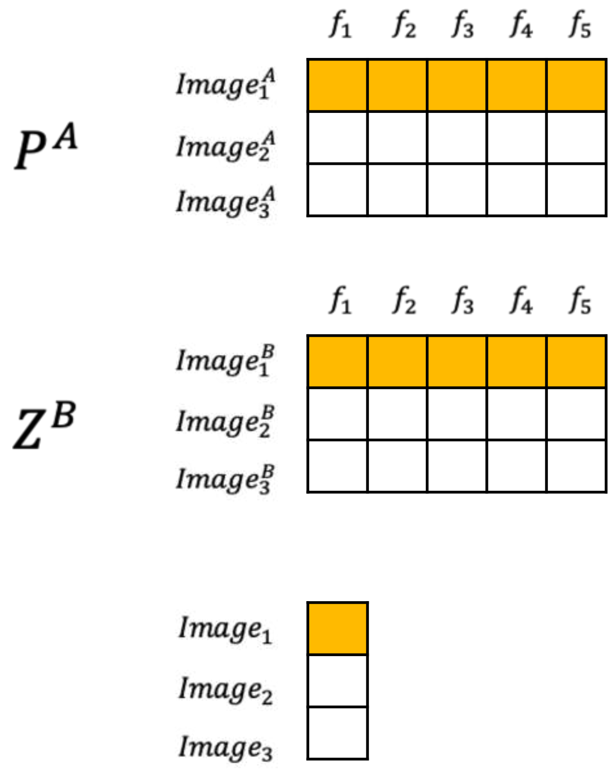
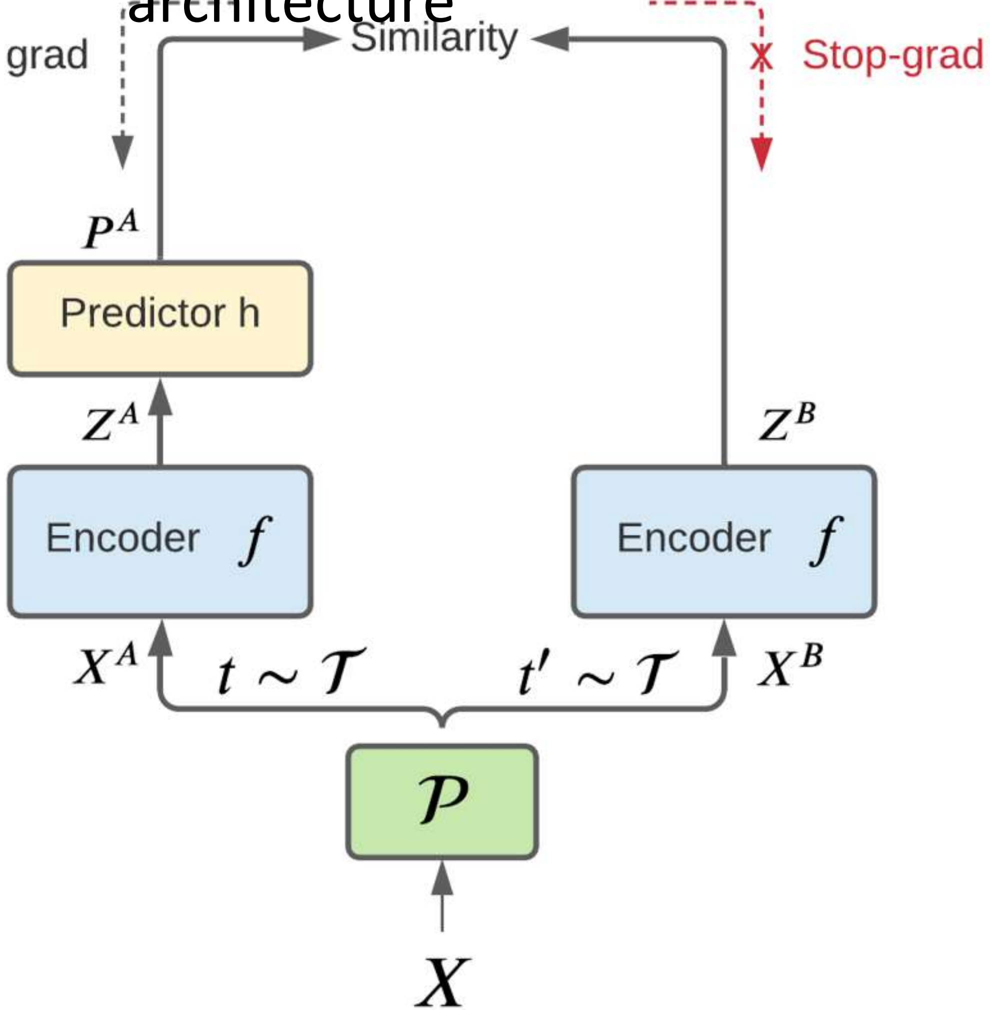


Image Representation Contrasting

Remember: in the SimSiam architecture



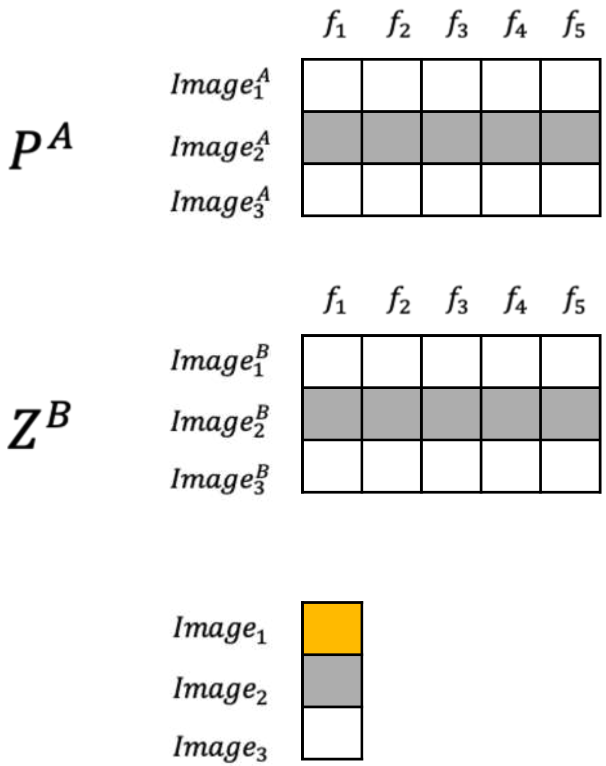
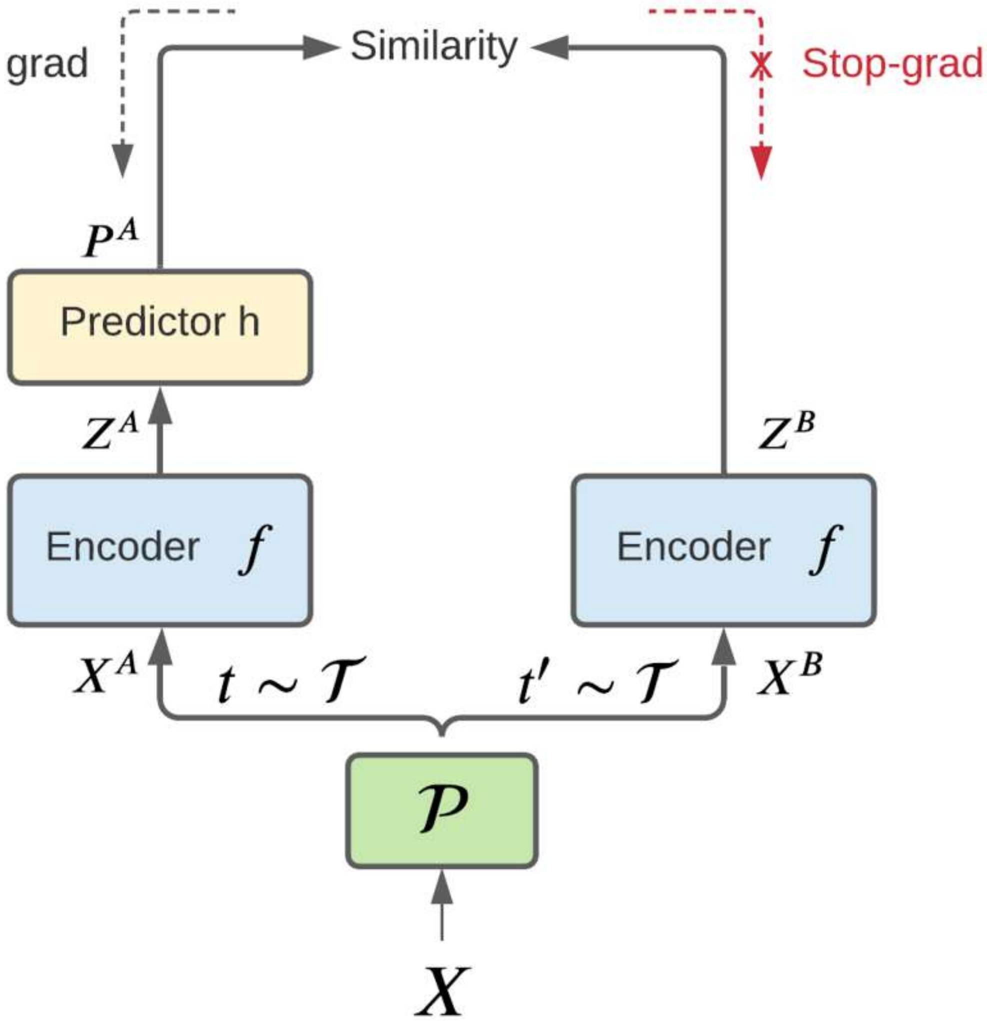


Image Representation Contrasting

Remember: in the SimSiam architecture



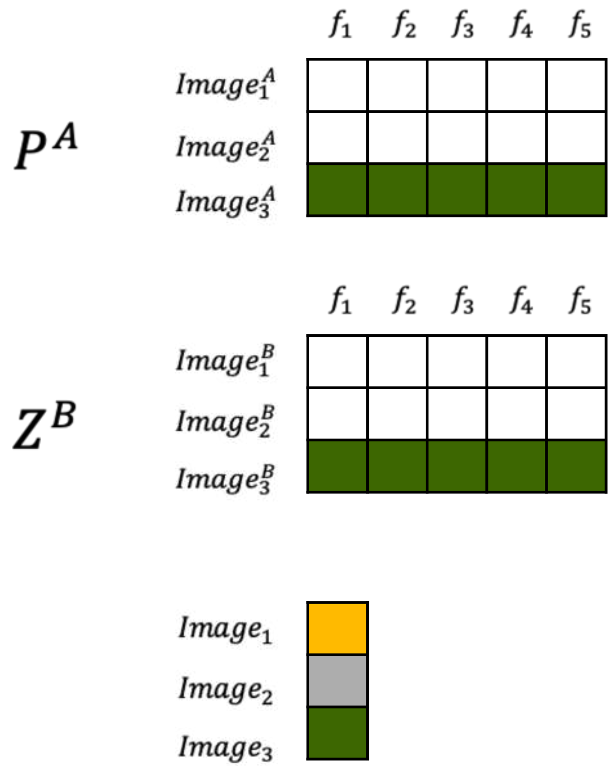
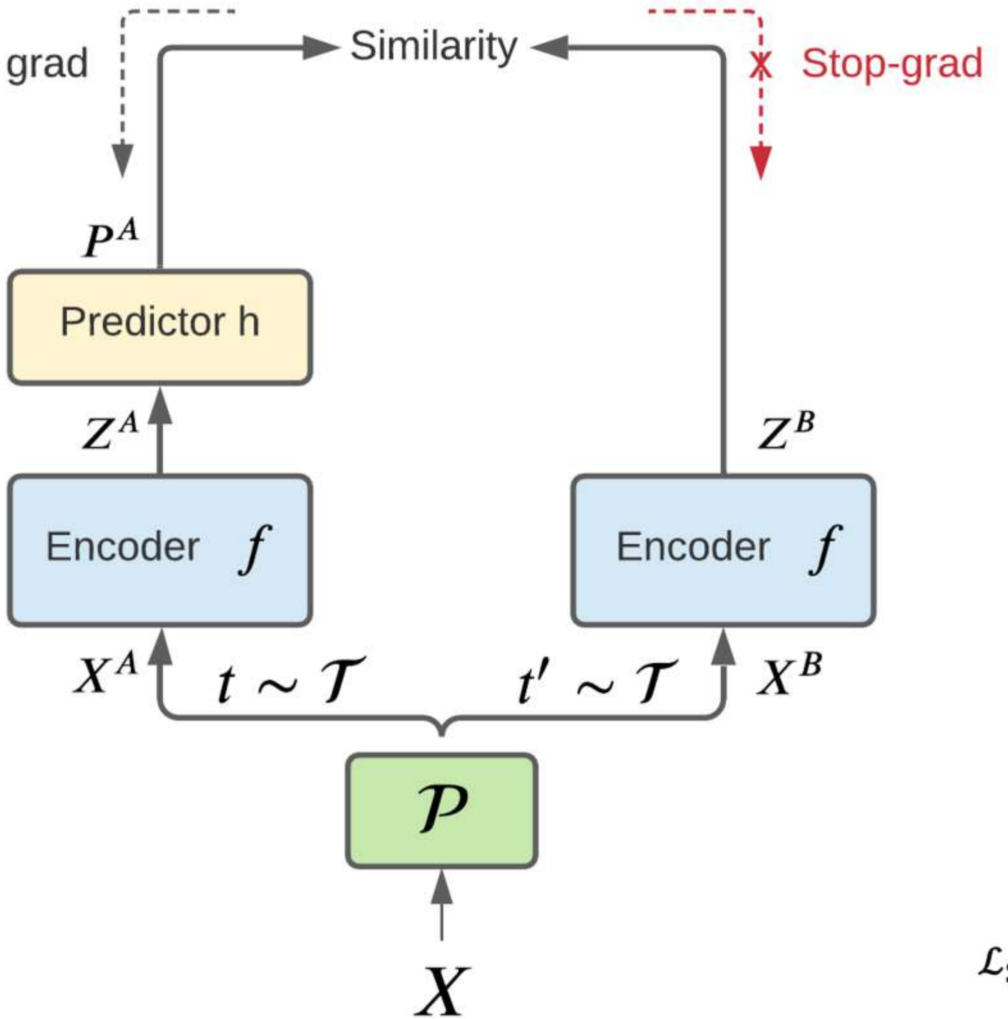
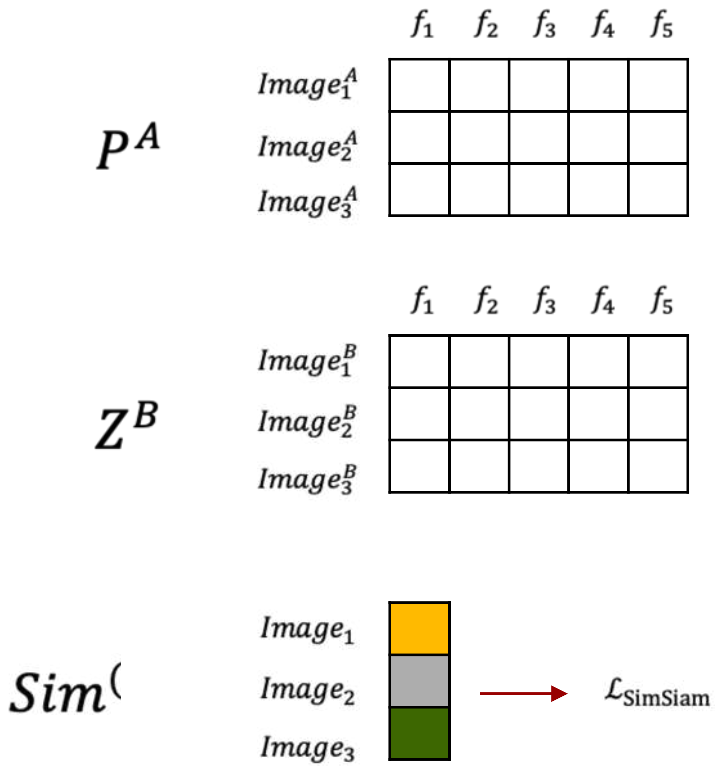


Image Representation Contrasting

SimSiam architecture





$$\mathcal{L}_{\text{SimSiam}} = -\frac{1}{2N}\sum_{i=1}^{N}\left(\left(Sim_i^{(1)}\right)^2 + \left(Sim_i^{(2)}\right)^2\right)$$

SimSiam

P^A

	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

P^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

Z^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

Z^A

	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

SimSiam

P^A

	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

P^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

Z^B

	f_1	f_2	f_3	f_4	f_5
$Image_1^B$					
$Image_2^B$					
$Image_3^B$					

Z^A

	f_1	f_2	f_3	f_4	f_5
$Image_1^A$					
$Image_2^A$					
$Image_3^A$					

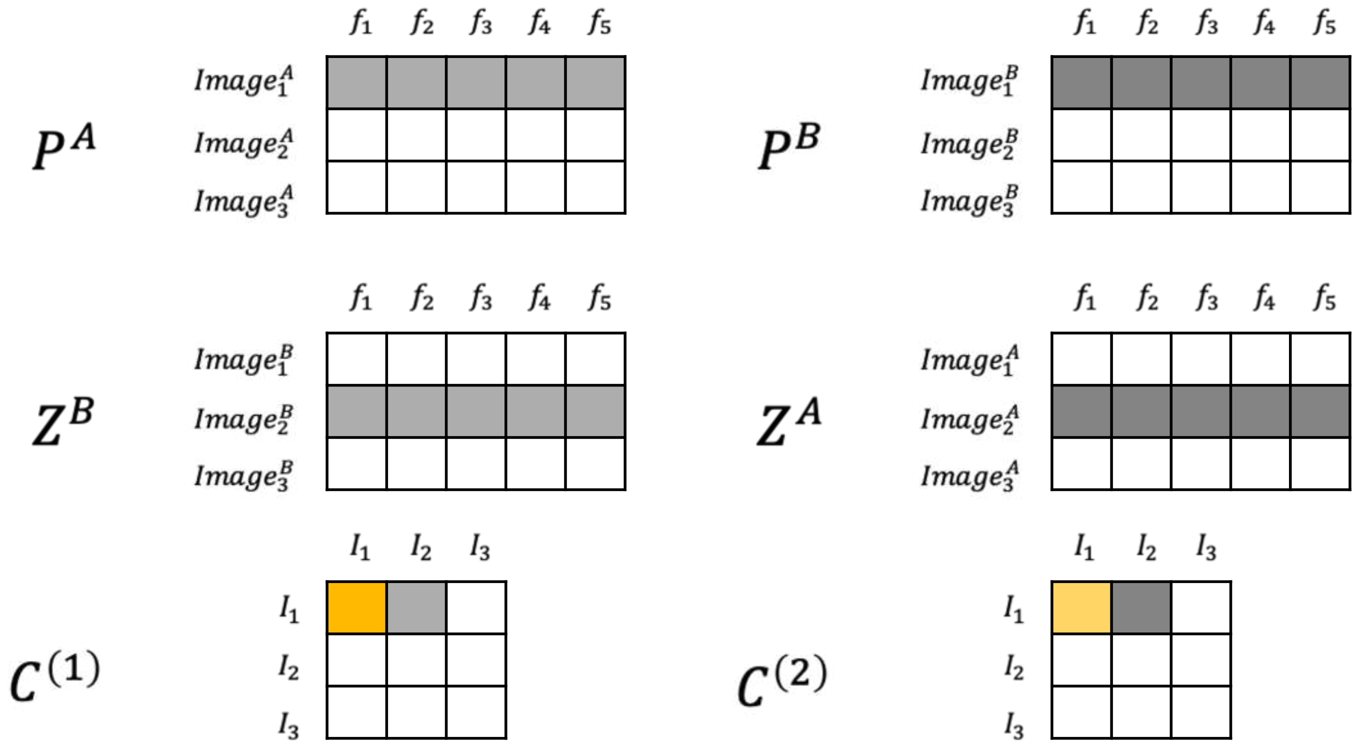
$C^{(1)}$

	I_1	I_2	I_3
I_1			
I_2			
I_3			

$C^{(2)}$

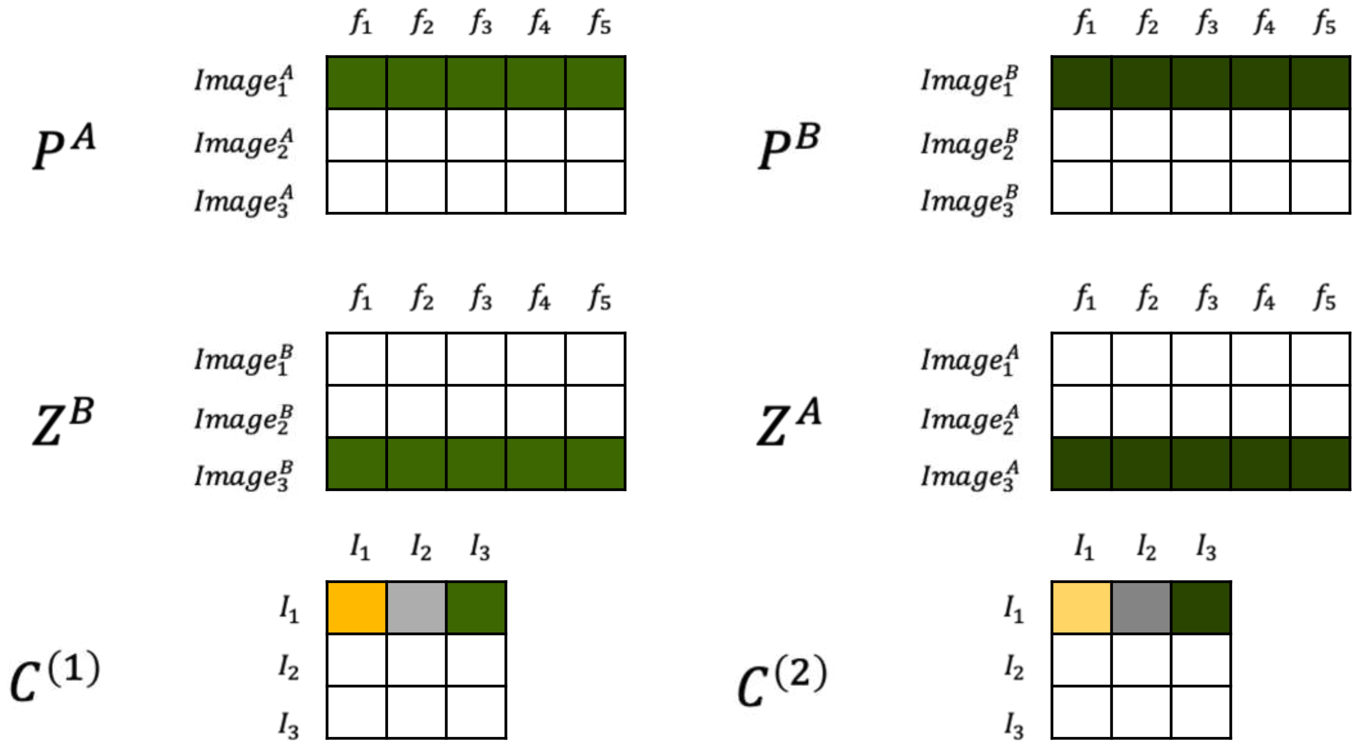
	I_1	I_2	I_3
I_1			
I_2			
I_3			

SimSiam



P^A
f_1 f_2 f_3 f_4 f_5
$Image_1^A$
$Image_2^A$
$Image_3^A$
P^B
$Image_1^B$
$Image_2^B$
$Image_3^B$
Z^B
Z^A
$C^{(1)}$
$C^{(2)}$
I_1 I_2 I_3

SimSiam



P^A
$Image_1^A$ $Image_2^A$ $Image_3^A$
f_1 f_2 f_3 f_4 f_5
P^B
$Image_1^B$ $Image_2^B$ $Image_3^B$
f_1 f_2 f_3 f_4 f_5
Z^B
$Image_1^B$ $Image_2^B$ $Image_3^B$
f_1 f_2 f_3 f_4 f_5
Z^A
$Image_1^A$ $Image_2^A$ $Image_3^A$
f_1 f_2 f_3 f_4 f_5
$C^{(1)}$
I_1 I_2 I_3
$C^{(2)}$
I_1 I_2 I_3

Our proposal 1: OptSSL

$$C^{(1)} \quad \begin{array}{c|c|c|c|} & I_1 & I_2 & I_3 \\ \hline I_1 & & & \\ \hline I_2 & & & \\ \hline I_3 & & & \\ \hline \end{array} \qquad\qquad C^{(2)} \quad \begin{array}{c|c|c|c|} & I_1 & I_2 & I_3 \\ \hline I_1 & & & \\ \hline I_2 & & & \\ \hline I_3 & & & \\ \hline \end{array}$$

OptSSL's Loss Function

$$\mathcal{L}_{Opt\text{-}SSL} = \mathcal{L}_{i\text{-}diag} + \lambda_1 \cdot \mathcal{L}_{i\text{-}off\,f\text{-}diag} + \mathcal{L}_{f\text{-}diag} + \lambda_2 \cdot \mathcal{L}_{f\text{-}off\,f\text{-}diag}$$

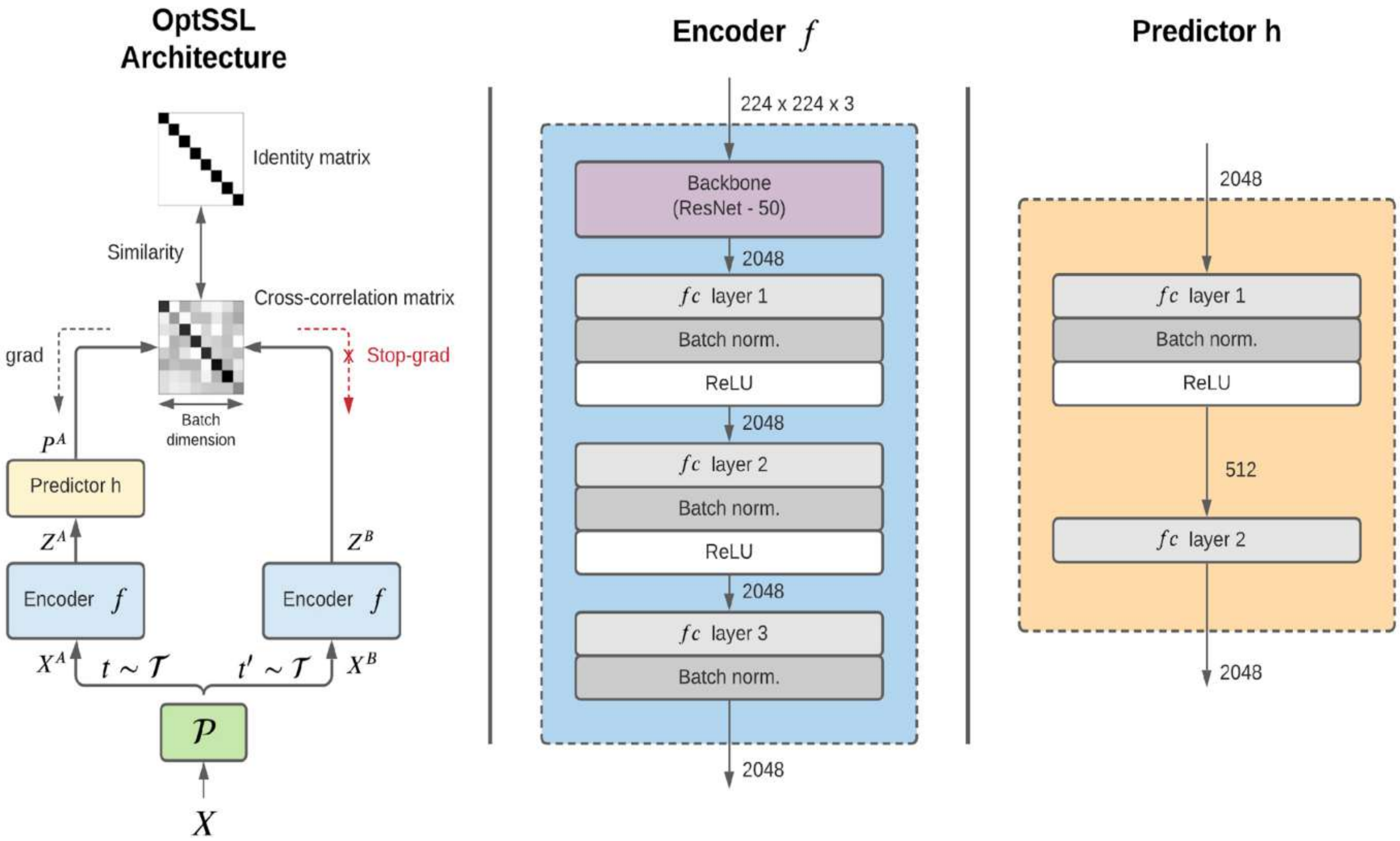
where

$$\mathcal{L}_{diag} = \sqrt{\frac{1}{2N}\left(\sum_{i=1}^{N}\left(1 - c_{ii}^{(1)}\right)^2 + \sum_{i=1}^{N}\left(1 - c_{ii}^{(2)}\right)^2\right)}$$

$$\mathcal{L}_{off-diag} = \sqrt{\frac{1}{2N(N-1)}\left(\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left(c_{ij}^{(1)}\right)^2 + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left(c_{ij}^{(2)}\right)^2\right)}$$

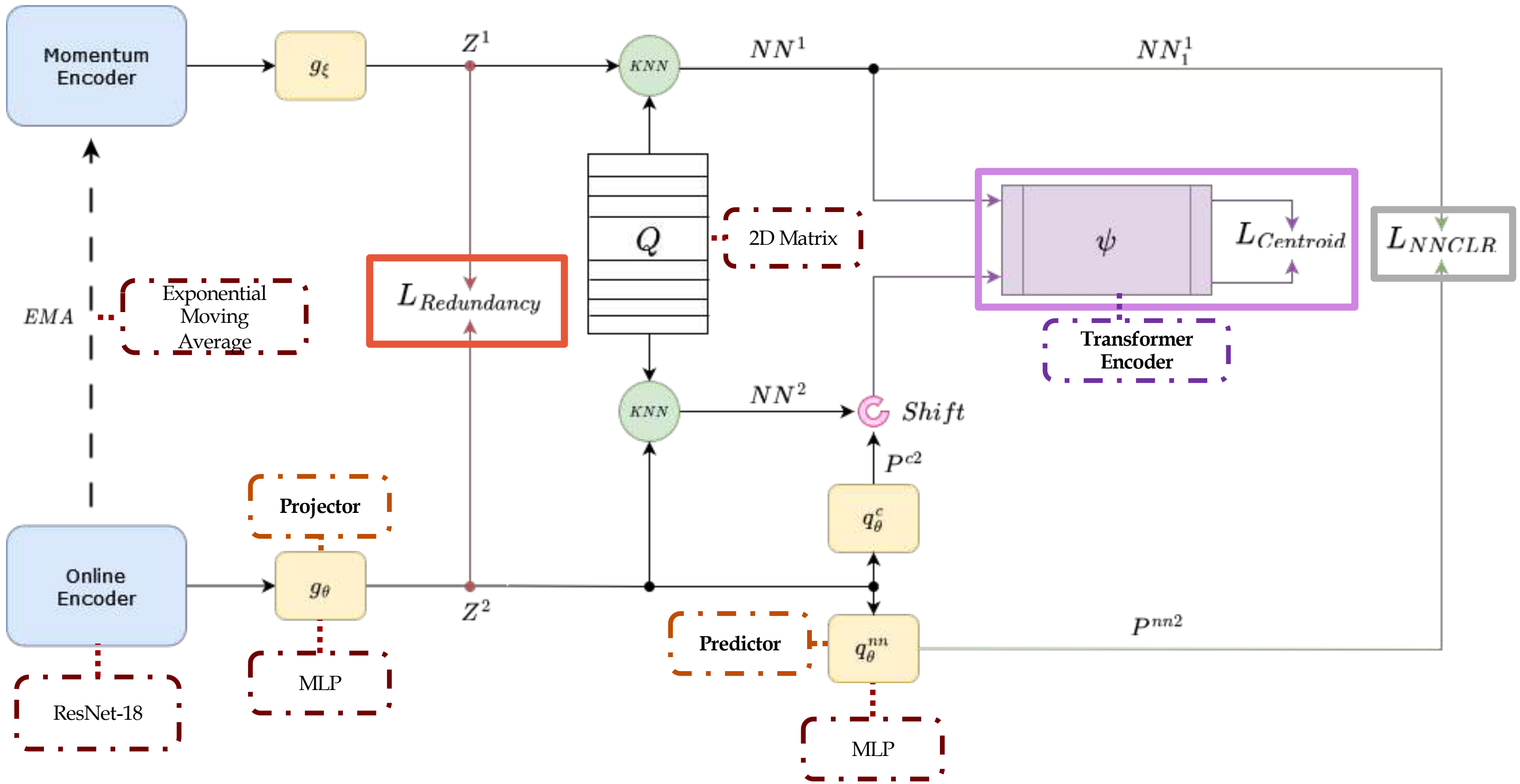
Applied both: to images in the batch and the features!

Our proposal 1: OptSSL

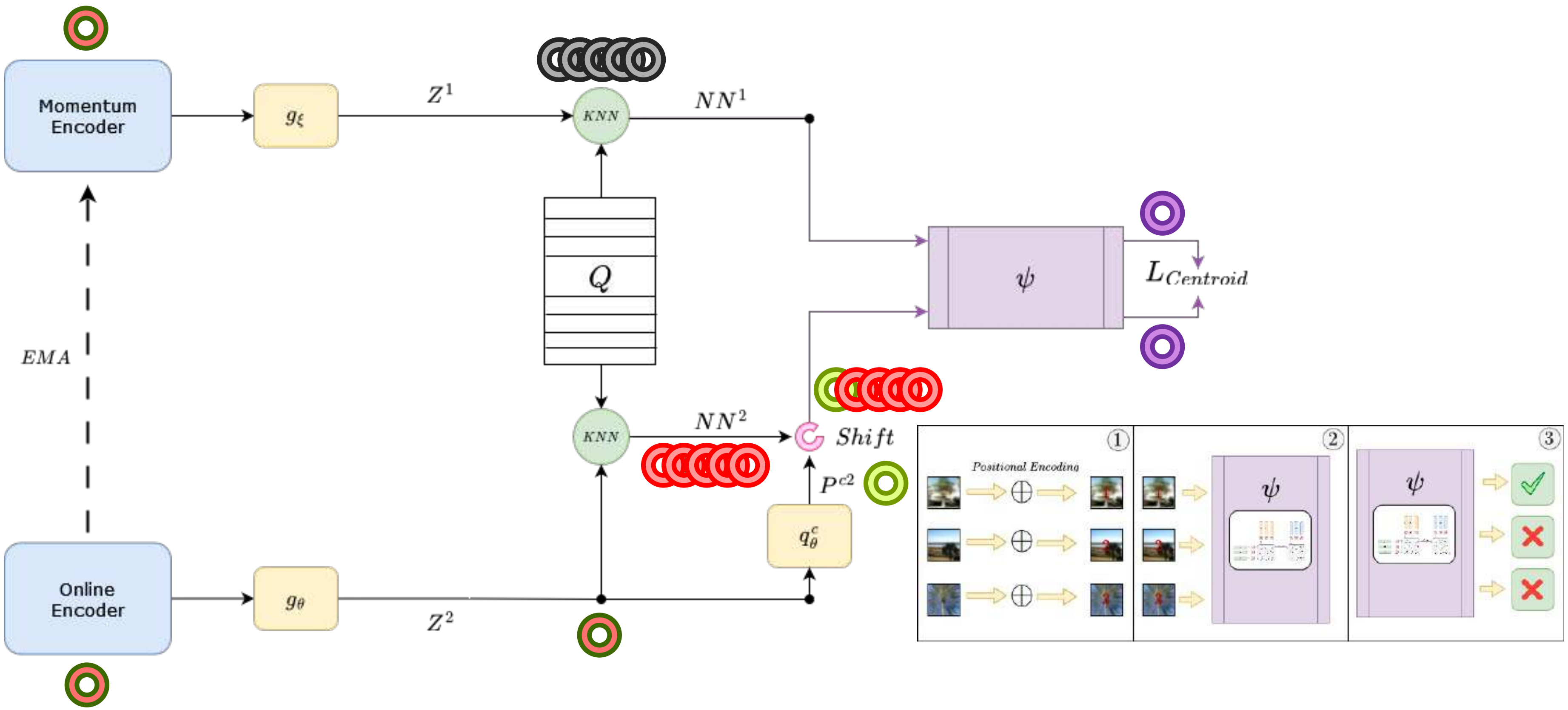


Nil Ballús, Bhalaji Nagarajan, Petia Radeva: Opt-SSL: An Enhanced Self-Supervised Framework for Food Recognition. IbPRIA 2022: 655-666

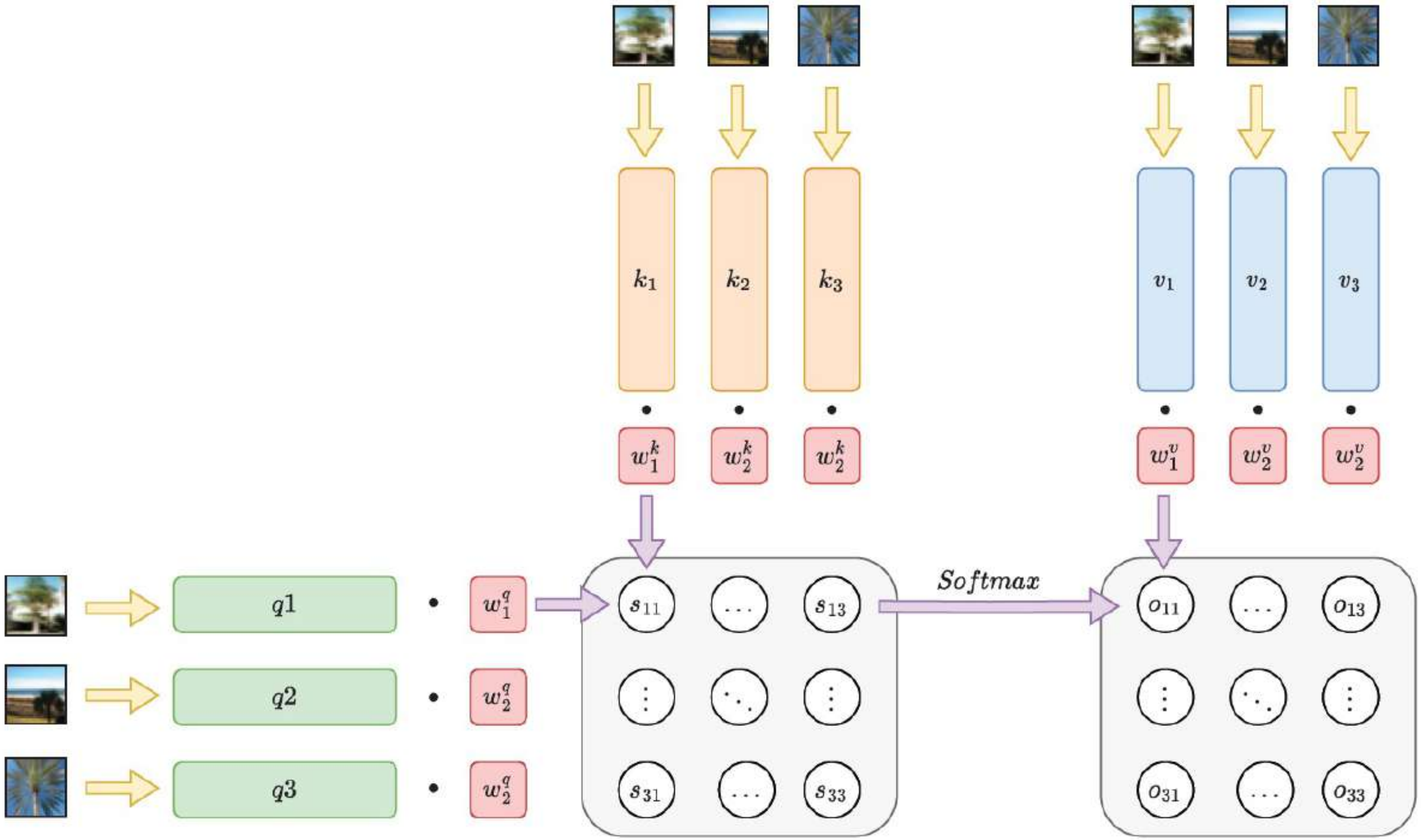
Our Proposal 2: Musketeer



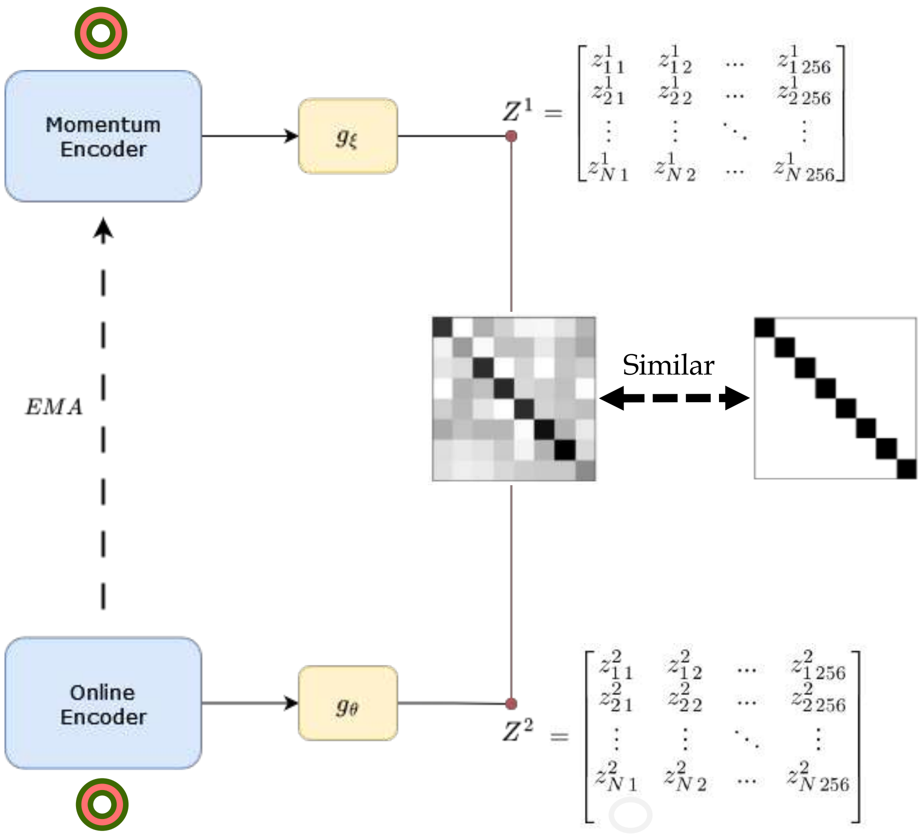
All for One: Centroid contrast



Self-Attention in order to obtain "context-aware" representations



Reducing Redundancy: Feature Contrast



Validation

Dataset and Evaluation Metrics

Dataset

Food-101

Evaluation Metrics

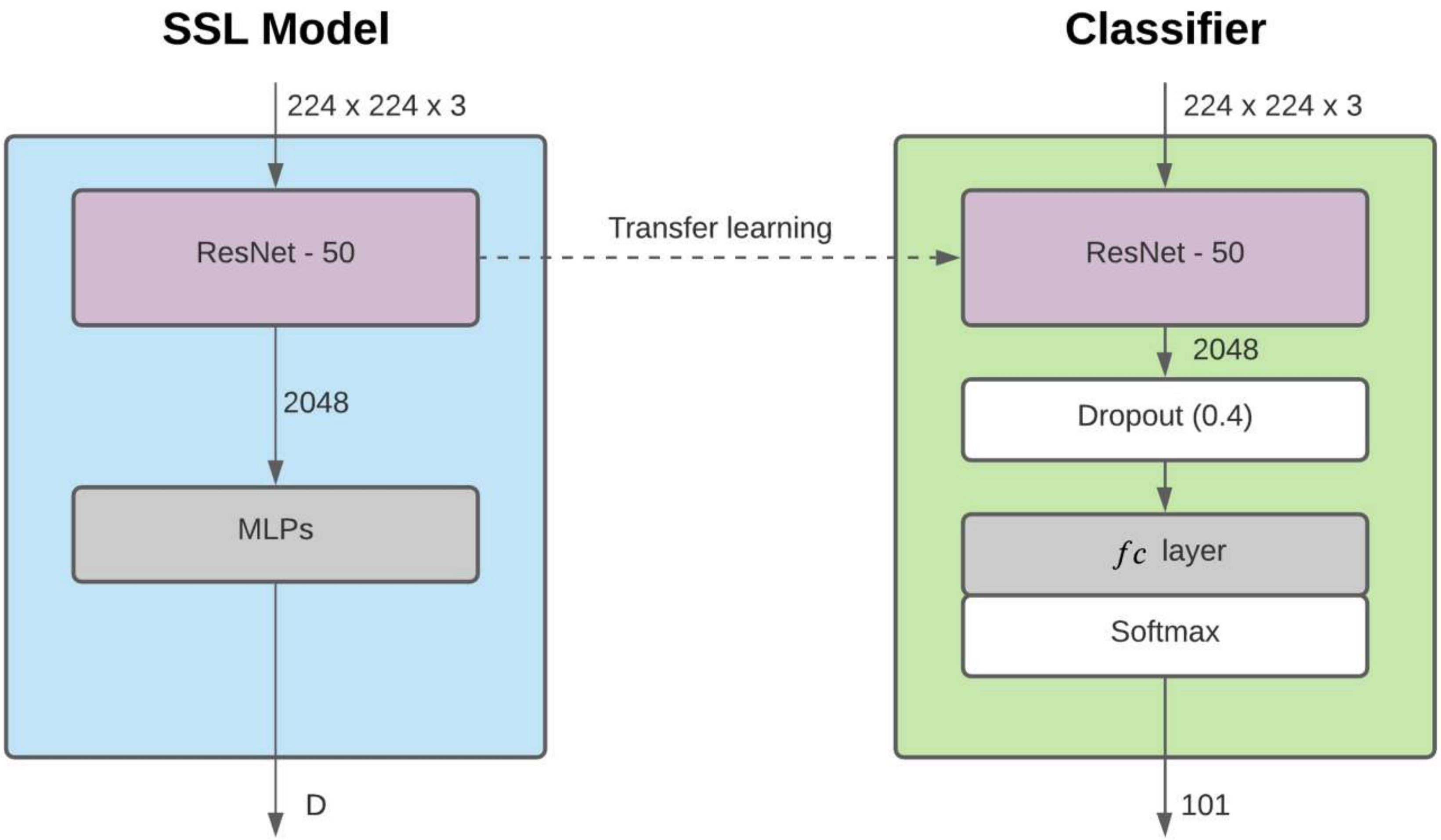
SSL Model

- Top 1 and top 5 accuracy using a k-NN classifier

Classifier

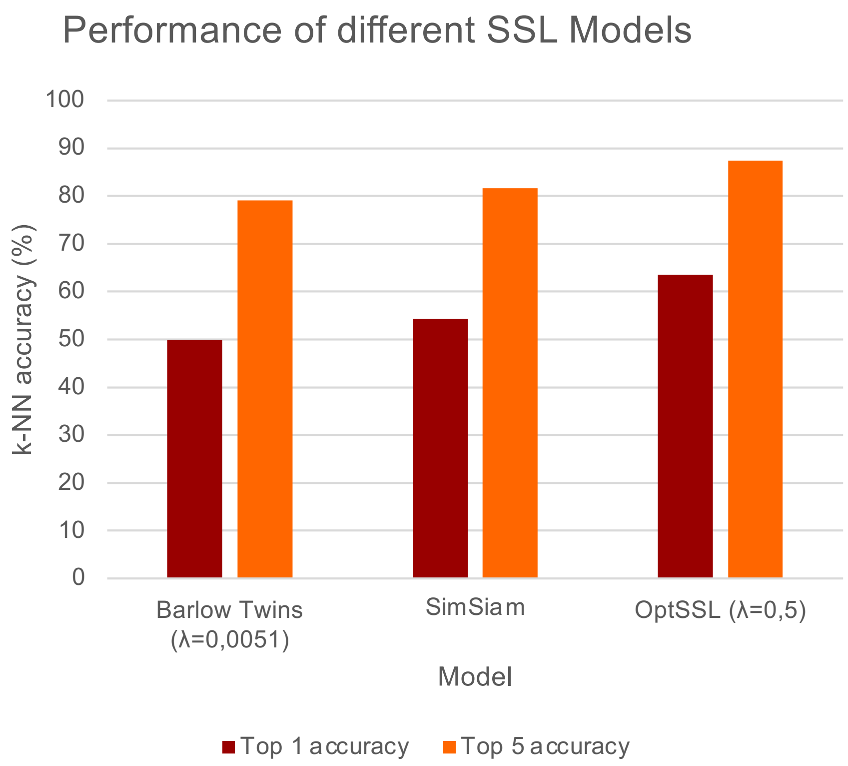
- Overall accuracy
- Variance
- Entropy
- Mutual Information

Baselines



SSL Model
Classifier
224 x 224 x 3
224 x 224 x 3
ResNet - 50
Transfer learning
ResNet - 50
2048
2048
Dropout (0.4)
MLPs
fc layer
Softmax
D
101

Results and analysis (SSL)



SSL Model	λ	K-NN Acc.	
		Top 1 acc. (%)	Top 5 acc. (%)
Barlow Twins	0,0051	49,89	79,07
SimSiam	-	54,33	81,62
OptSSL	0,5	**63,52**	**87,48**

What about Musketeer?

SSL Frameworks: Results

Framework	Main Characteristics
SimCLR	Starting point, use of Negative samples, high batch size, contrastive loss
BYOL	Randomly initialized network instead of Negative samples, MSE loss, lower reliance on batch size, introduces the predictor
SimSiam	SimCLR without Negative samples, BYOL without the target network, use of stop-gradient, siamese structure
Barlow Twins	Innovative loss function that uses cross-correlation matrixes, use of batch normalization, benefits from high dimensional representations
NNCLR	Introduces NN algorithm to provide more richness, modified version of SimCLR

Method	Top-1	Top-5
PIRL	63.6	-
CPC v2	63.8	85.3
PCL	65.9	-
CMC	66.2	87.0
MoCo v2	71.1	-
SimSiam	71.3	-
SimCLR v2	71.7	-
SwAV	71.8	-
InfoMin Aug.	73.0	91.1
BYOL	74.3	91.6
NNCLR (ours)	**75.4**	**92.3**
BARLOW TWINS (ours)	73.2	91.0

ImageNet linear classification results from Debidatta Dwibedi et al. (2021)

Quantitative Results: CIFAR

CIFAR-10

Method	Backbone	Epochs	Acc@1 (Online)	Acc@5 (Online)
BYOL	ResNet18	1000	92.58	99.79
DeepCluster V2	ResNet18	1000	88.85	99.58
DINO	ResNet18	1000	89.52	99.71
MoCo V2+	ResNet18	1000	92.94	99.79
MoCo V3	ResNet18	1000	93.10	99.80
ReSSL	ResNet18	1000	90.63	99.62
SimCLR	ResNet18	1000	90.74	99.75
Simsiam	ResNet18	1000	90.51	99.72
SwAV	ResNet18	1000	89.17	99.68
VIbCReg	ResNet18	1000	91.18	99.74
VICReg	ResNet18	1000	92.07	99.74
W-MSE	ResNet18	1000	88.67	99.68
Barlow Twins	ResNet18	1000	92.10	99.73
NNCLR	ResNet18	1000	91.88	99.78
Musketeer (Ours)	ResNet18	1000	**93.24**	**99.88**

CIFAR-100

Method	Backbone	Epochs	Acc@1 (Online)	Acc@5 (Online)	k-NN Acc@1 (Online)
BYOL	ResNet18	1000	70.46	91.96	-
DeepCluster V2	ResNet18	1000	63.61	88.09	-
DINO	ResNet18	1000	66.76	90.34	-
MoCo V2+	ResNet18	1000	69.89	91.65	-
MoCo V3	ResNet18	1000	68.83	90.57	-
ReSSL	ResNet18	1000	65.92	89.73	-
SimCLR	ResNet18	1000	65.78	89.04	-
Simsiam	ResNet18	1000	66.04	89.62	-
SwAV	ResNet18	1000	64.88	88.78	-
VIbCReg	ResNet18	1000	67.37	90.07	-
VICReg	ResNet18	1000	68.54	90.83	-
W-MSE	ResNet18	1000	61.33	87.26	-
NNCLR	ResNet18	1000	69.62	91.52	-
NNCLR*	ResNet18	1000	69.17	91.70	62.16
Barlow Twins	ResNet18	1000	70.90	91.91	-
Barlow Twins*	ResNet18	1000	71.21	92.46	63.11
MSF*	ResNet18	1000	67.84	91.64	63.36
Musketeer (Ours)	ResNet18	1000	**72.17**	**93.35**	**64.84**

Quantitative Results: ImageNet-100

Method	Backbone	Epochs	Acc@1 (online)	Acc@5 (online)
BYOL ++	ResNet18	400	80.16	95.02
DeepCluster V2	ResNet18	400	75.36	93.22
DINO	ResNet18	400	74.84	92.92
MoCo V2+ ++	ResNet18	400	78.20	95.50
MoCo V3 ++	ResNet18	400	80.36	95.18
ReSSL	ResNet18	400	76.92	94.20
SimCLR ++	ResNet18	400	77.64	94.06
Simsiam	ResNet18	400	74.54	93.16
SwAV	ResNet18	400	74.04	92.70
VIbCReg	ResNet18	400	79.86	94.98
VICReg ++	ResNet18	400	79.22	95.06
W-MSE	ResNet18	400	67.60	90.94
Barlow Twins ++	ResNet18	400	80.38	95.28
NNCLR ++	ResNet18	400	79.80	95.28
Musketeer (Ours)	ResNet18	400	**81.93**	**96.23**

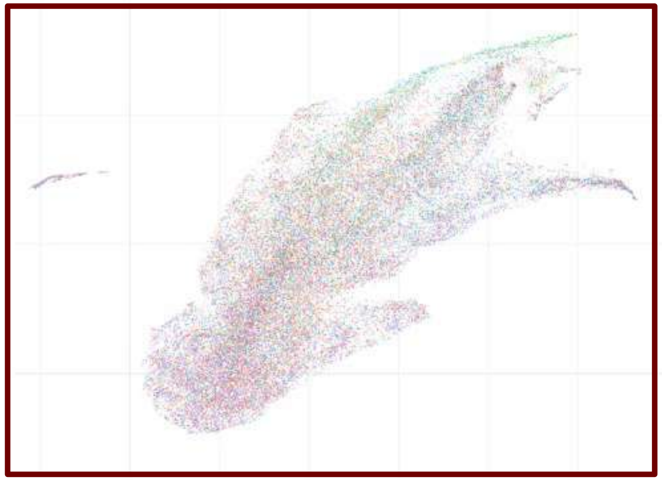
Quantitative Results: Objective importance

Method	NNCLR	Centroid	Redundancy	EMA	Acc@1	NN Acc@1
Musketeer (v0)	✓	✗	✗	✗	69.62	68.8
Musketeer (v1)	✗	✓	✗	✗	67.4	82.8
Musketeer (v2)	✓	✓	✗	✗	71.02	85.28
Musketeer (v3)	✓	✓	✗	✓	71.08	**86.16**
Musketeer (v4)	✗	✓	✓	✓	71.31	80.6
Musketeer (v5)	✓	✗	✓	✓	71.64	78.8
Musketeer (v6)	✓	✓	✓	✓	**72.17**	82.16

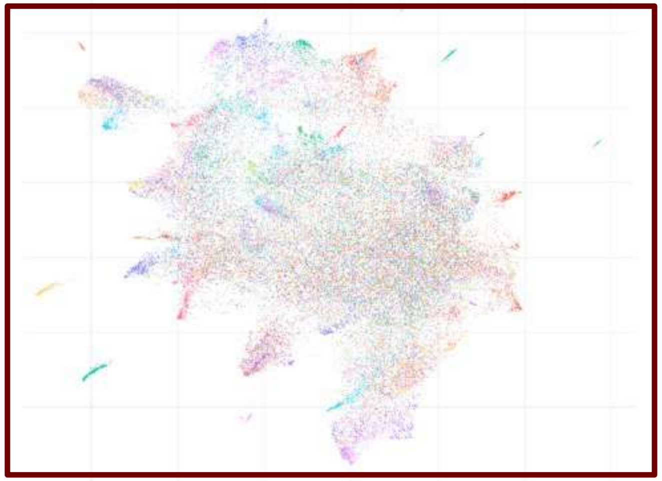
✓ = Included ✗ = Not included

Qualitative Analysis: UMAP

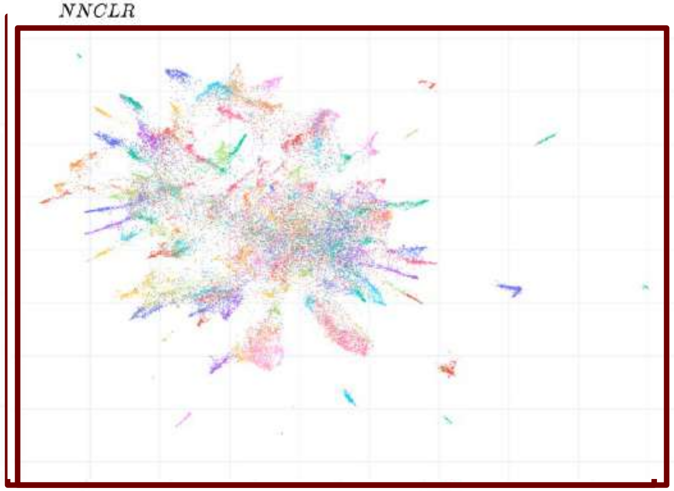
Epoch 1



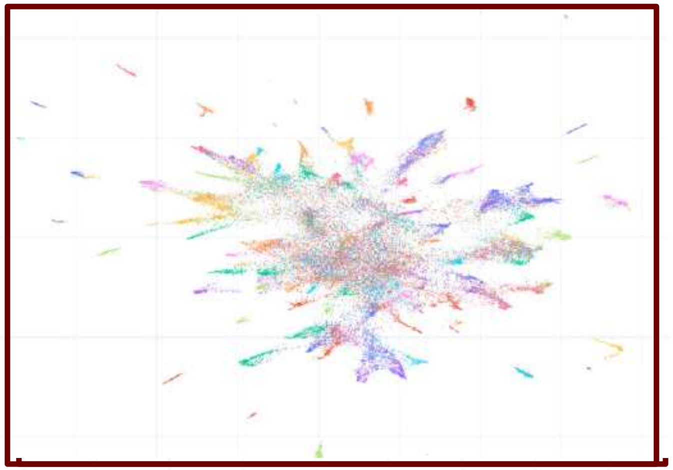
Epoch 100



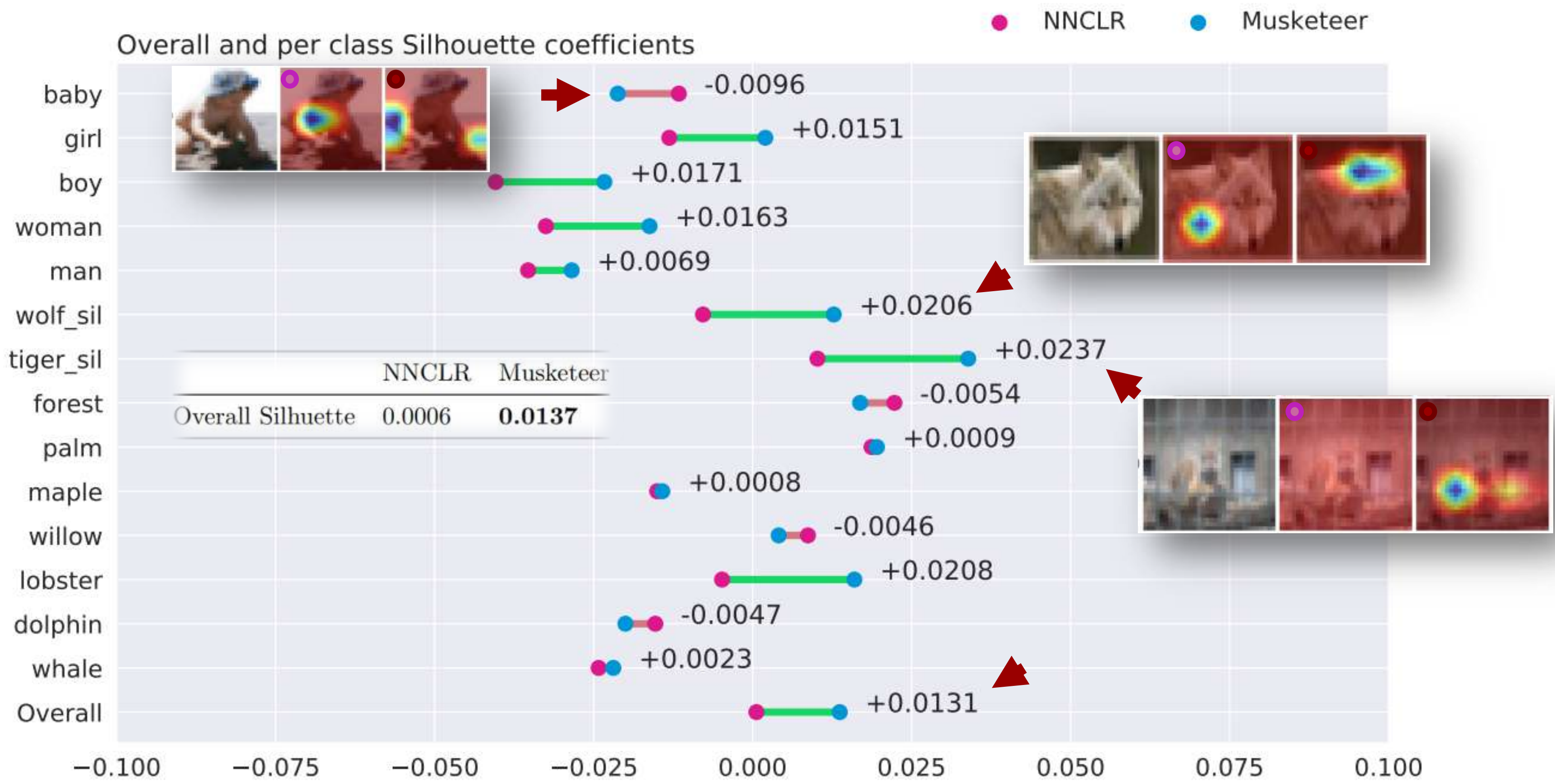
NNCLR



Musketeer

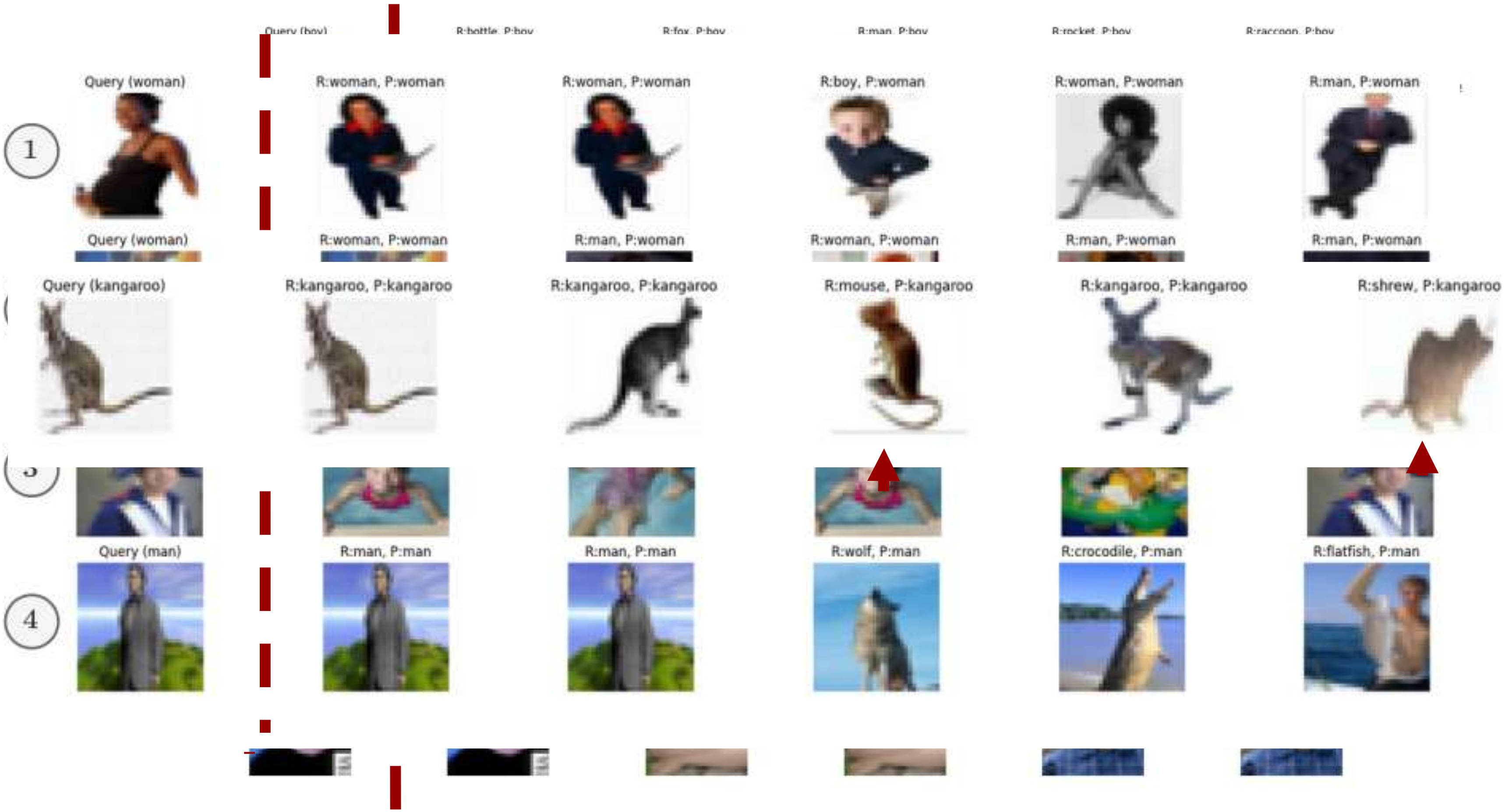


Qualitative Analysis: Silhouette



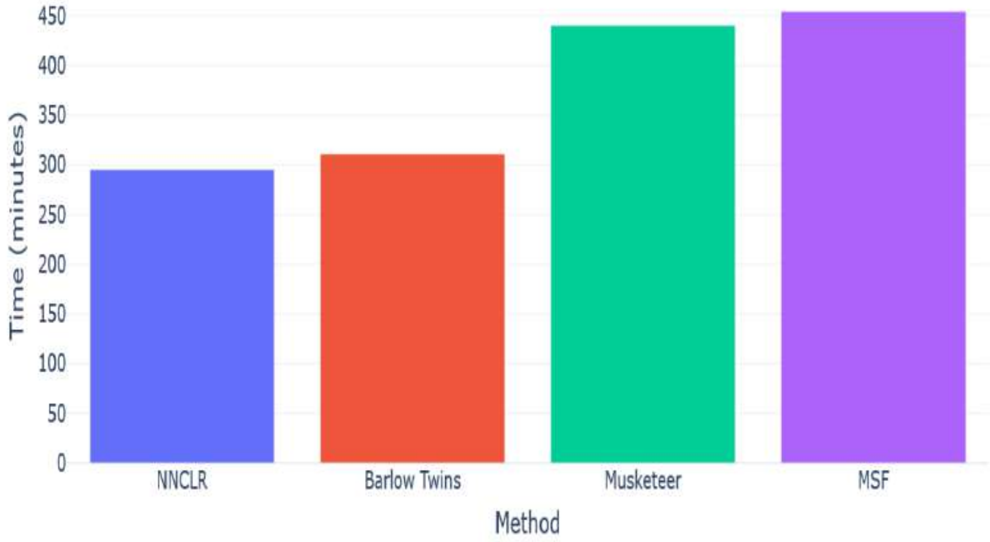
	NNCLR	Musketeer
Overall Silhuette	0.0006	**0.0137**

Qualitative Analysis: NN Retrieval



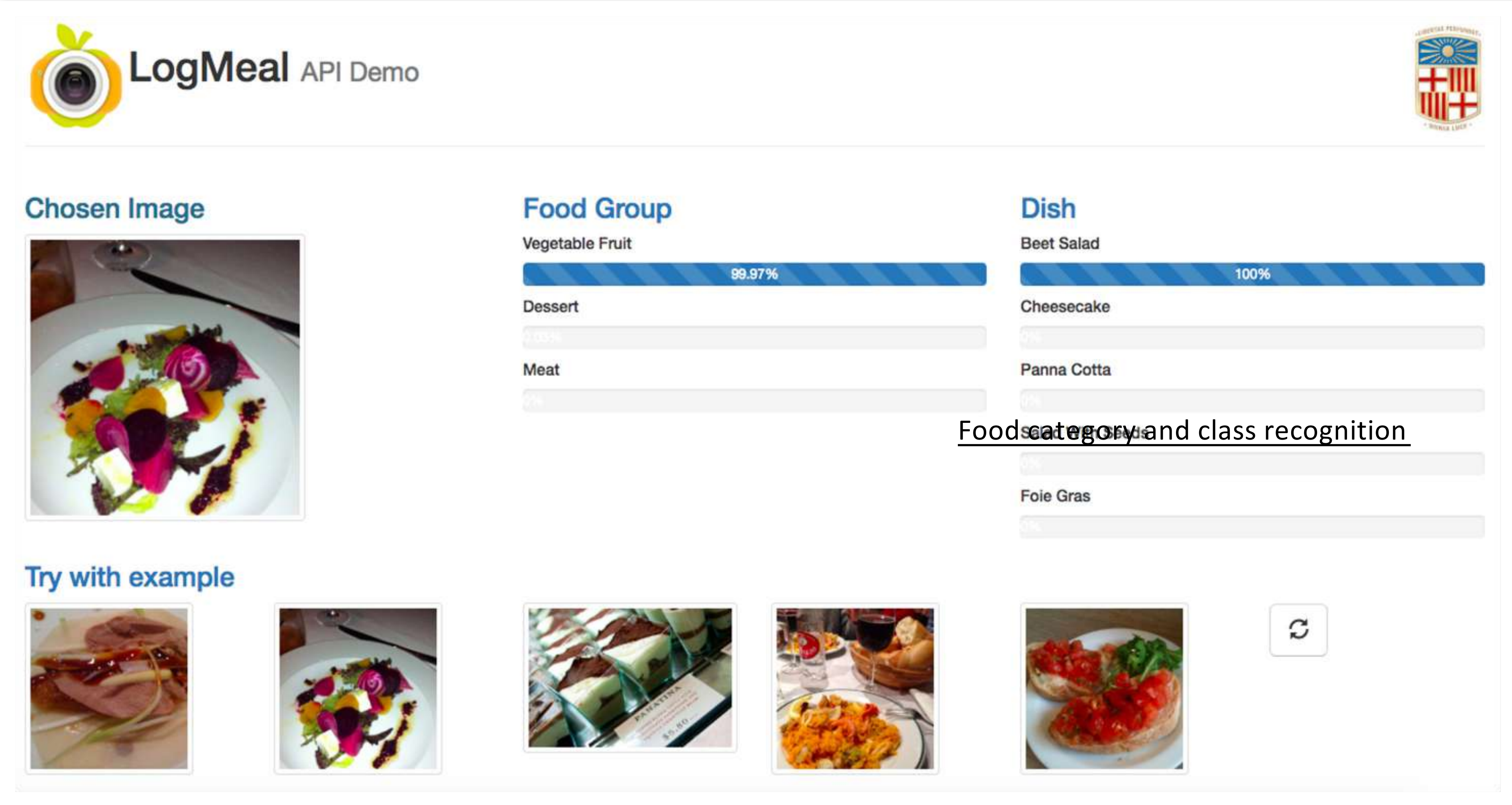
Features of Musketeer

- Not very sensitive regarding the number of neighbours extracted.
- More expensive than single neighbour contrast.



Speaking about Food Applications

Food recognition

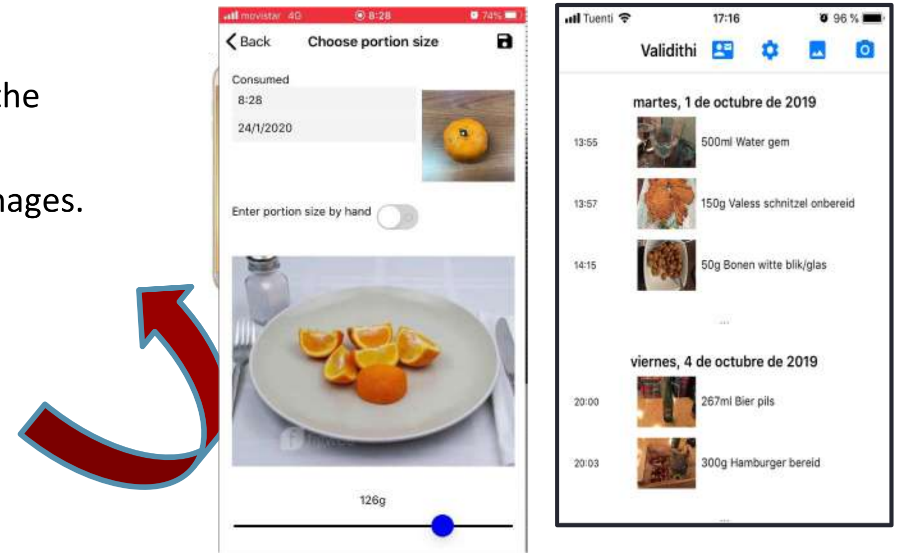


Food category and class recognition

Try it: www.logmeal.es/demo

SUCCESS STORY: FOOD intake monitoring of kidney transplant patients

LogMeal is a HealthApp and API in the cloud that is able to automatically recognize and analyze food from images.

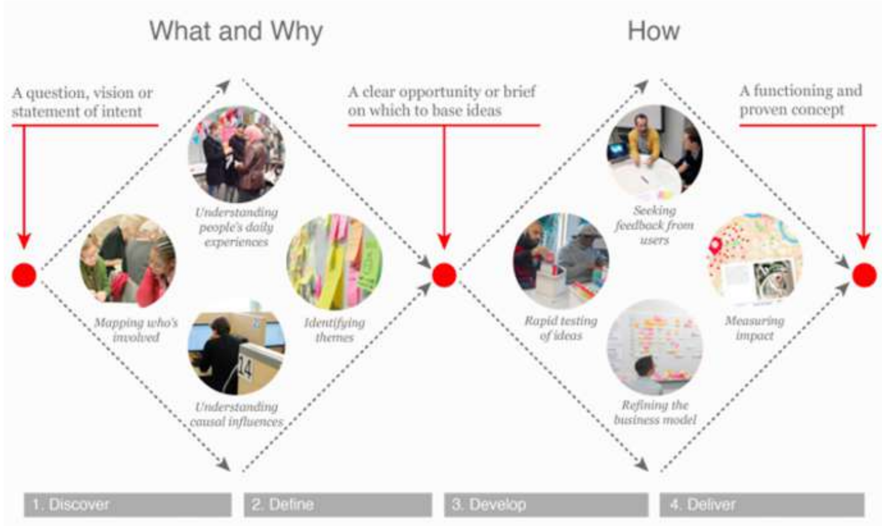


Validithi (EIT Health, 2019/20)

- **Automatic food diary** construction (UB).
- **Accurate, objective and continuous** food intake monitoring (UB).
- Semi-automatic **volume estimation** (Nestle).
- **Meal planner** and **health recommendations (**Nestle**)**.

21:22

SUCCESS STORY: FOOD intake monitoring for malnutrition prevention in elderly



NESTORE developed a multi-dimensional, personalized coaching system to support healthy ageing:

1) Generating and sustaining motivation to take care of health;

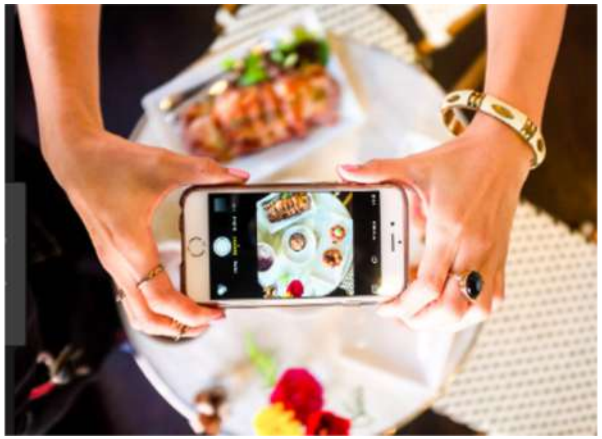
1) Suggesting healthy nutrition and personalized physical and mental coach, as well as social interaction, to prevent decline and preserve wellbeing.

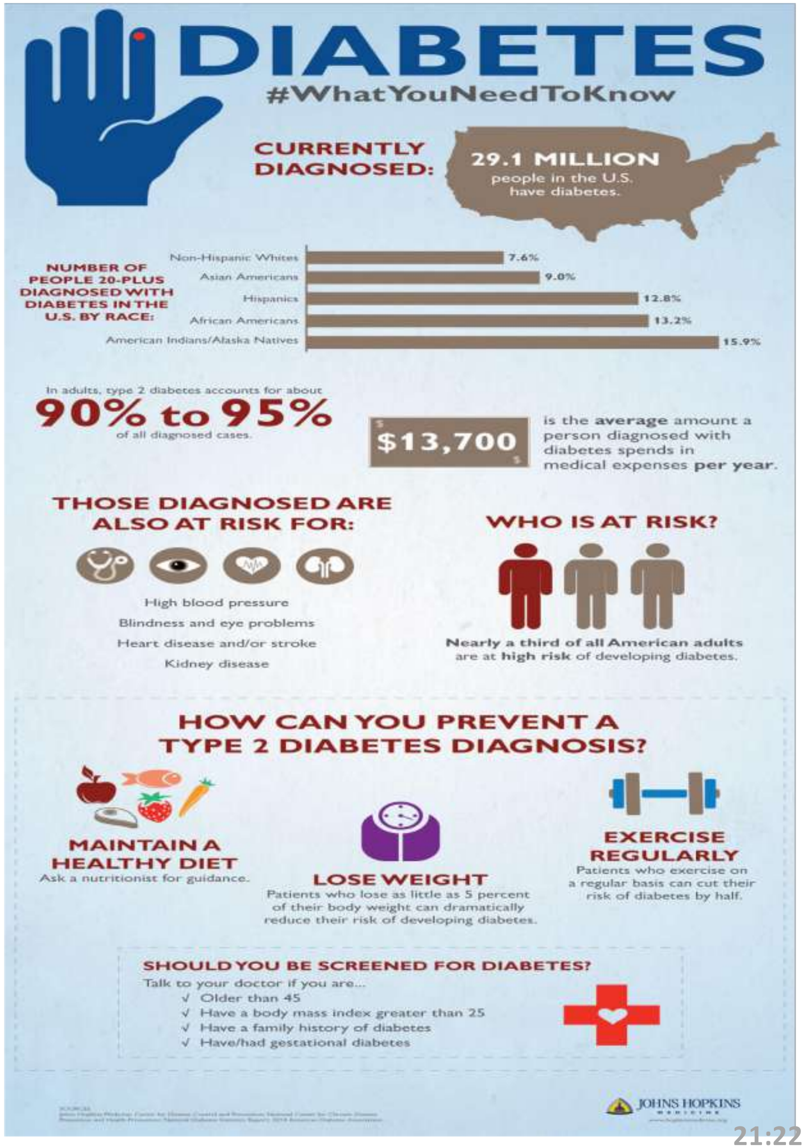
POLITECNICO DI MILANO · TU Delft · UNIVERSITAT DE BARCELONA · eurecat Centre Tecnològic de Catalunya · HE SO · RO PARDO · flex LIVE SMARTER

SUCCESS STORY Diacare: mHealth app to assist diabetic patients

The number of people with diabetes has increased from 108 million in 1980 to 422 million in 2017.

Pulso Edicions and UB are developing an app oriented to diabetic people in order to monitor their food intake and receive objective and timely feedback.





eurostars™

BYS GRUP

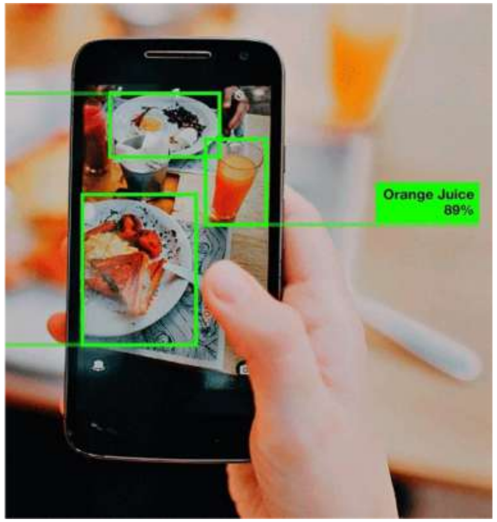
SUCCESS STORY Greenhabit: a serious game to promote change behaviour

Ted Talk of Chantal Linders: “Manage the Monster in Your Head”

Greenhabit (EIT Digital, 2020/21)

Success story: Aigecko Technologies

Touchless Checkout System: Plate and food recognition Multiple Payment System
User identification (NFC, QR, Face recognition, company card)



LogMeal API

API that allows food recognition (ready meals and food) using Artificial Intelligence algorithms with just a photo.

UNIVERSITAT DE BARCELONA

Noticias Agenda UBtv Sal

APRENDE INVESTIGA

Noticias

Inicio > Noticias > Nace AIGecko Technologies, la inteligencia artificial al servicio del reconocimiento...

Nace AIGecko Technologies, la inteligencia artificial al servicio del reconocimiento de imágenes

De izquierda a derecha, Marc Bolaños, Petia Radeva, M.ª Carme Verdaguer y Eric Verdaguer.

20/01/2021

Recerca

Ha nacido la nueva *spin-off* AIGecko, que ofrece servicios de reconocimiento y análisis de imágenes basados en los algoritmos desarrollados por miembros del Grupo de Investigación Computer Vision and Machine Learning de la UB, y que han dado pie a diferentes productos. La actividad de la empresa se enmarca en el aprendizaje profundo (*deep learning*), un campo de la inteligencia artificial que en los últimos años ha revolucionado el mercado y nuestro día a día en aspectos como la conducción automática, la visión artificial o los asistentes virtuales, entre muchas otras aplicaciones.

Más información

AIGecko's Food Image Analysis Applications

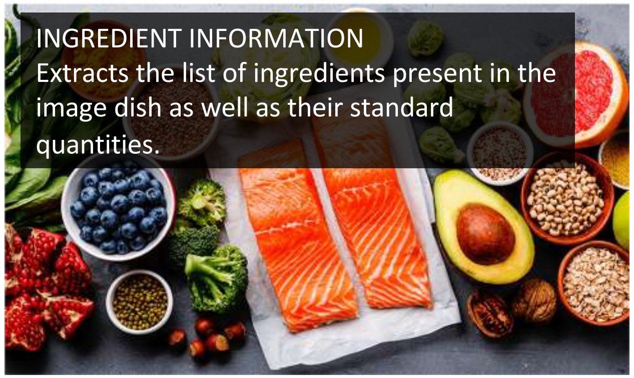
FOOD TYPE DETECTION
API to detect cooked food, prepared food, beverages, fresh vegetables and fruits, non-food products and more.

FOOD GROUP DETECTION
Detects the basic food groups present in food. Ideal for the generation of food records and food diaries.

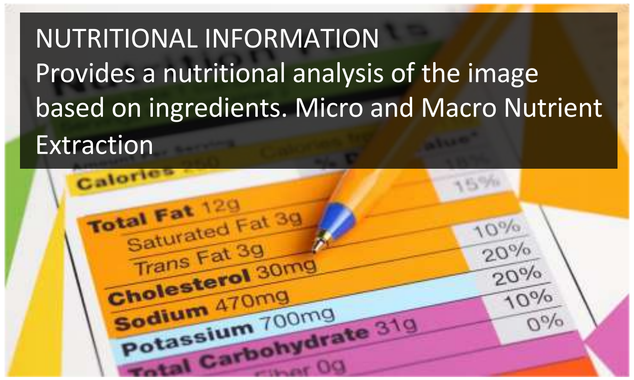
SINGLE DISH RECOGNITION
Detects more than 880 different local and international dishes from any cuisine in the world.

RECOGNITION OF VARIOUS DISHES
Recognises and lists all the foods present in a combination dish.

INGREDIENT INFORMATION
Extracts the list of ingredients present in the image dish as well as their standard quantities.



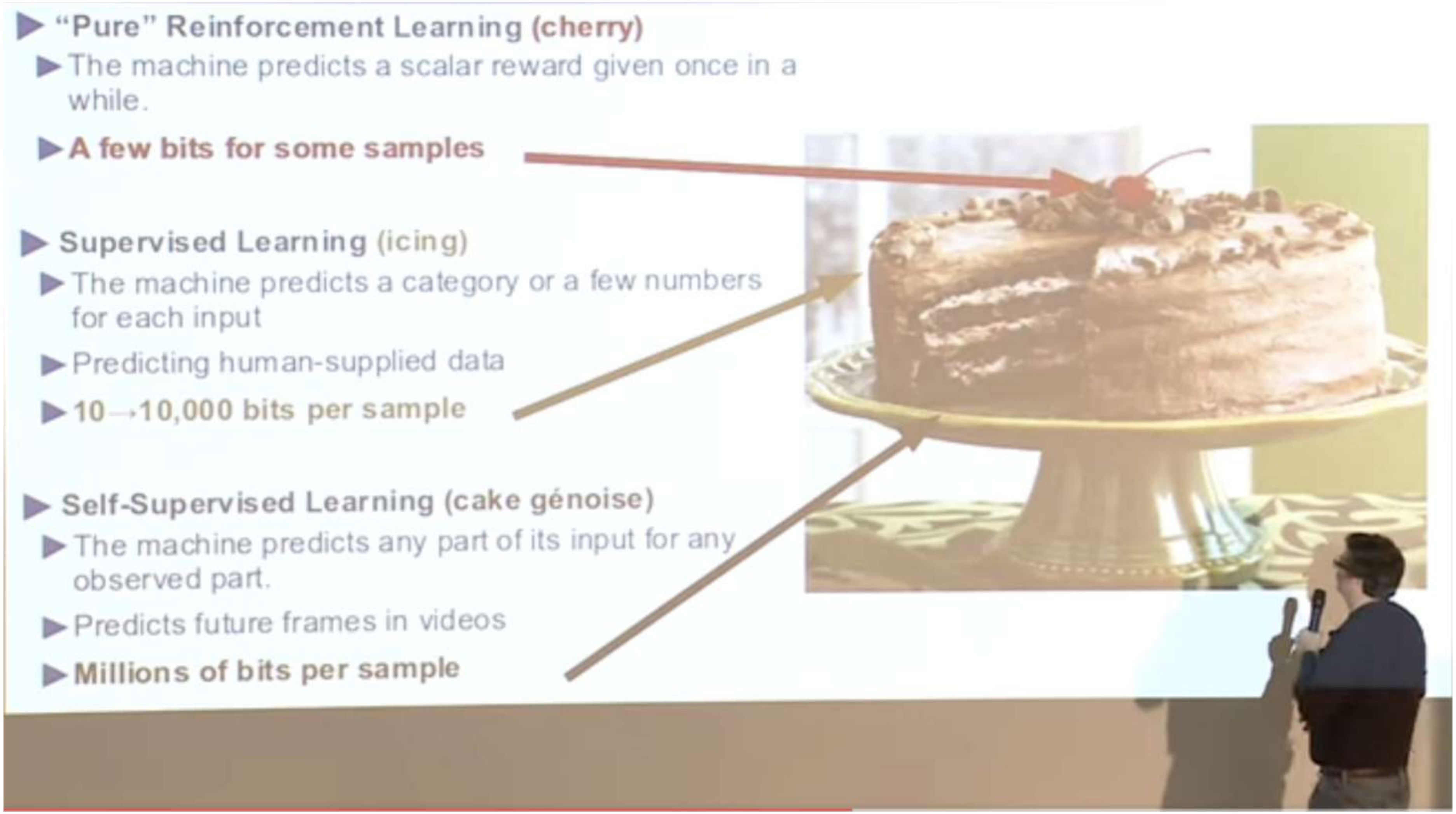
NUTRITIONAL INFORMATION
Provides a nutritional analysis of the image based on ingredients. Micro and Macro Nutrient Extraction



Conclusions

- OptSSL method **outperforms SimSiam and Barlow Twins** for the food image representation task.
  - Showed the importance of **contrasting both positive and negative** samples.
- Musketeer introduces **Self-attention operations to create single representations, defined as centroids**, from the extracted neighbours.
  - increases the neighbour retrieval accuracy while avoiding efficiency loss.
- Musketeer **combines its neighbour contrast objective with a feature redundancy reduction** objective, forming a symbiosis that proves to be beneficial in the overall performance of the framework.
- Musketeer **consistently outperforms SoTA** instance discrimination frameworks on popular image classification benchmarking datasets, namely, CIFAR-10, CIFAR-100 and ImageNet-100.
- **Food** Image Analysis is **highly underexplored problem** that could convert in an important **benchmark** for CV algorithms.
- Multiple **real applications** and **professional opportunities**

How much information is the Machine given during Learning?



21:22


Thank you!
UNIVERSITAT DE BARCELONA
CVC
Centre de Visió per Computador
ICPRAM 2023
12th International Conference on Pattern Recognition Applications and Methods