IDEAS NCBR

Divide and not forget: Ensemble of selectively trained experts in Continual Learning

Grzegorz Rypeść

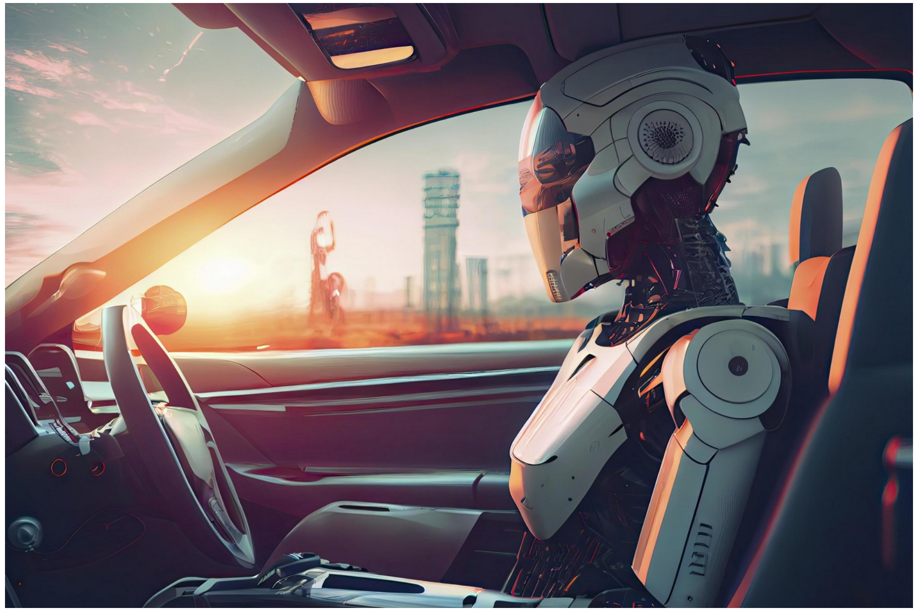
Agenda

- Continual learning
- How neural networks forget?
- Algorithm SEED

Continual learning

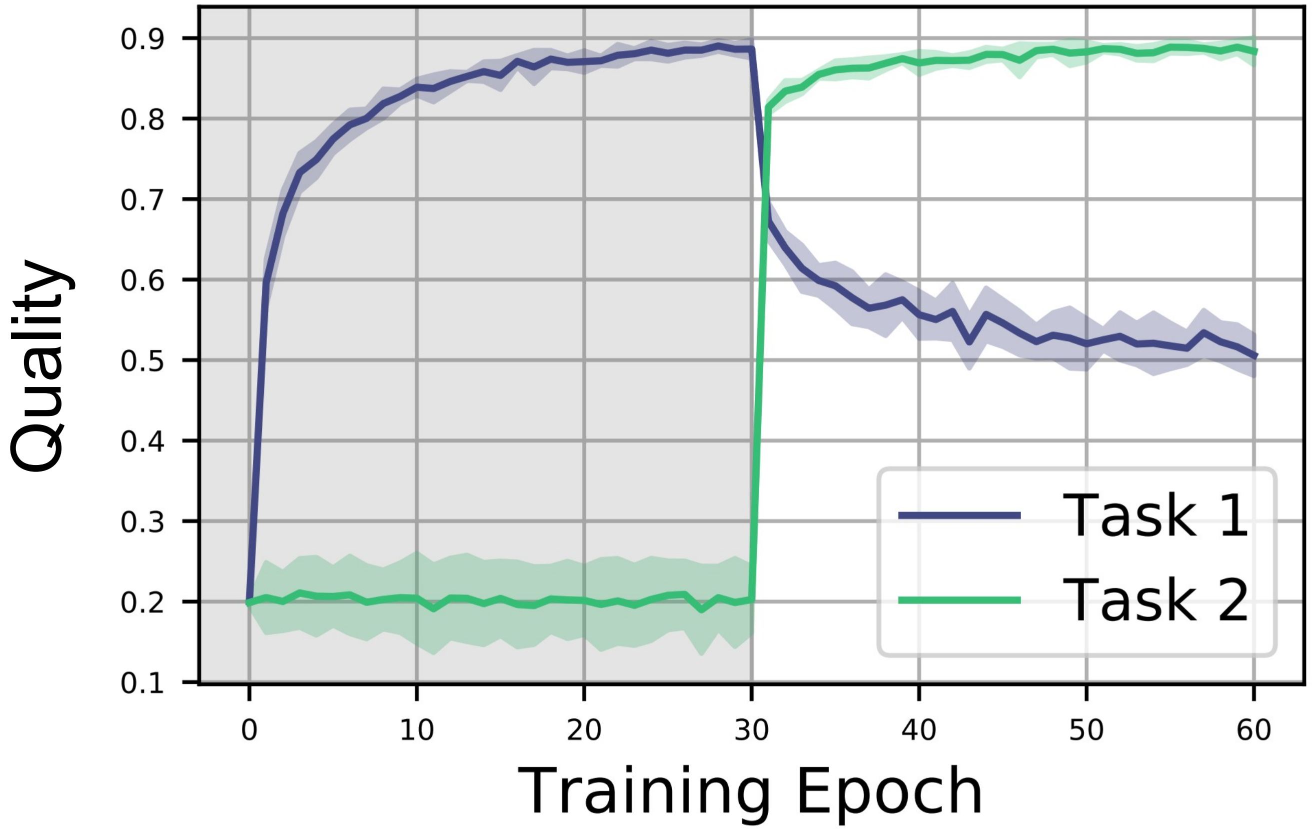
- Problem of training a model for a large number of sequential tasks without forgetting knowledge obtained from the preceding tasks
- Artificial neural networks suffer from catastrophic forgetting

???



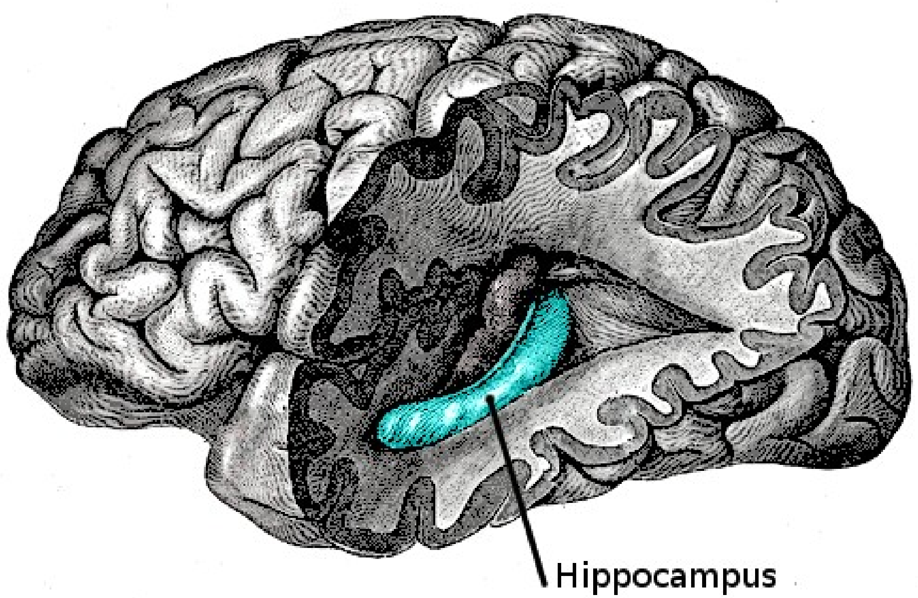
Continual learning

- Problem of learning a model for a large number of sequential tasks without forgetting knowledge obtained from the preceding tasks
- Artificial neural networks suffer from catastrophic forgetting



Continual learning

- Problem of learning a model for a large number of sequential tasks without forgetting knowledge obtained from the preceding tasks
- Artificial neural networks suffer from
- How did biology solve this problem?

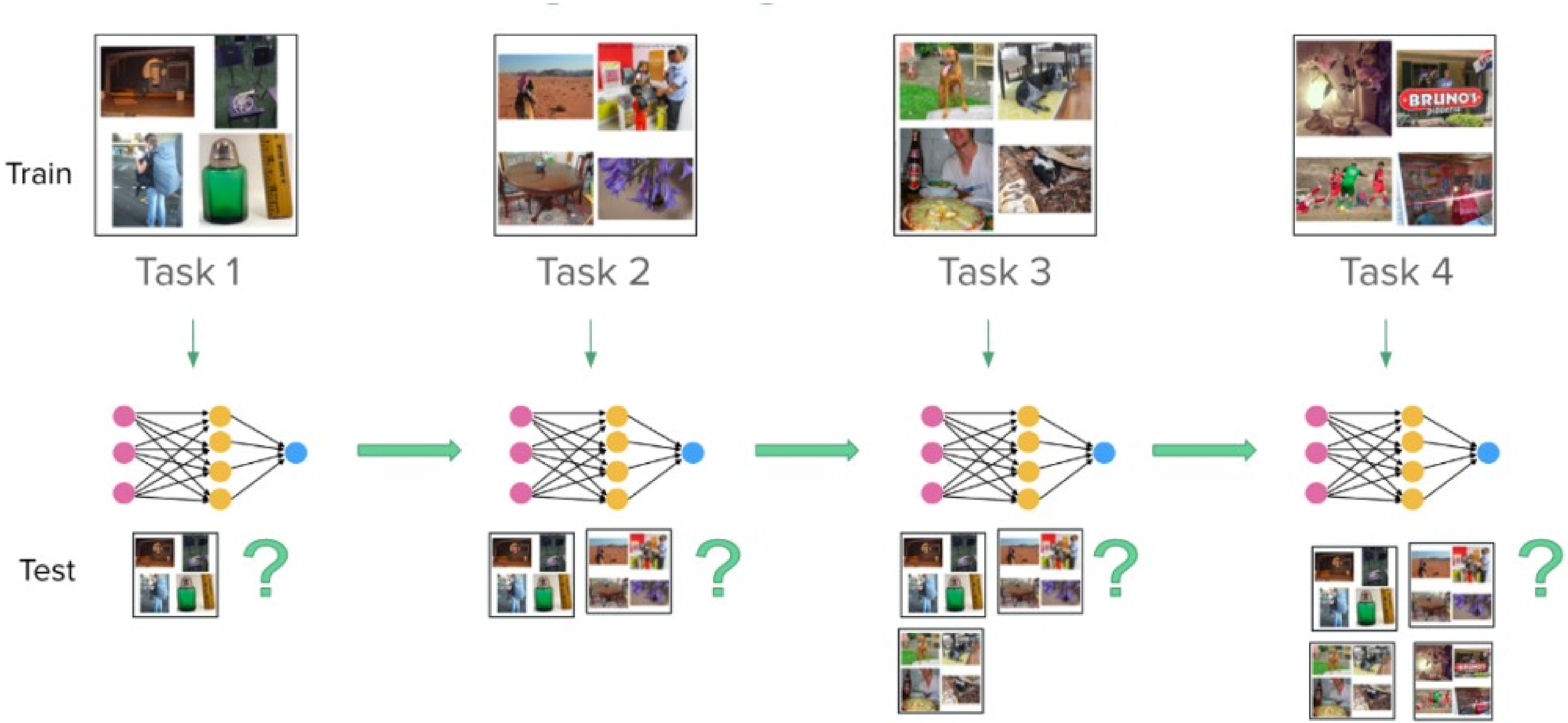


[1]

Continual learning - motivation

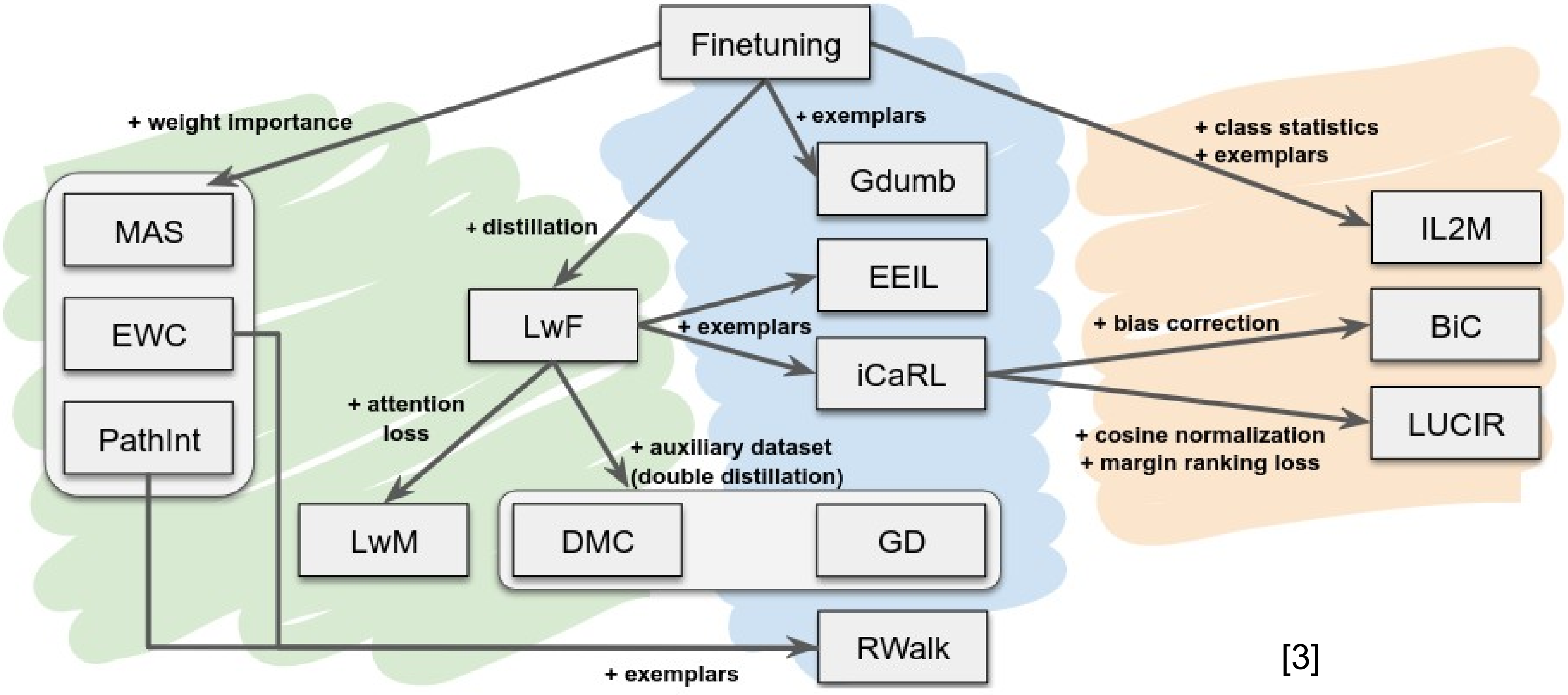
- Computational performance
- Ecology
- Privacy concerns
- General artificial intelligence (AGI)

Class Incremental Learning



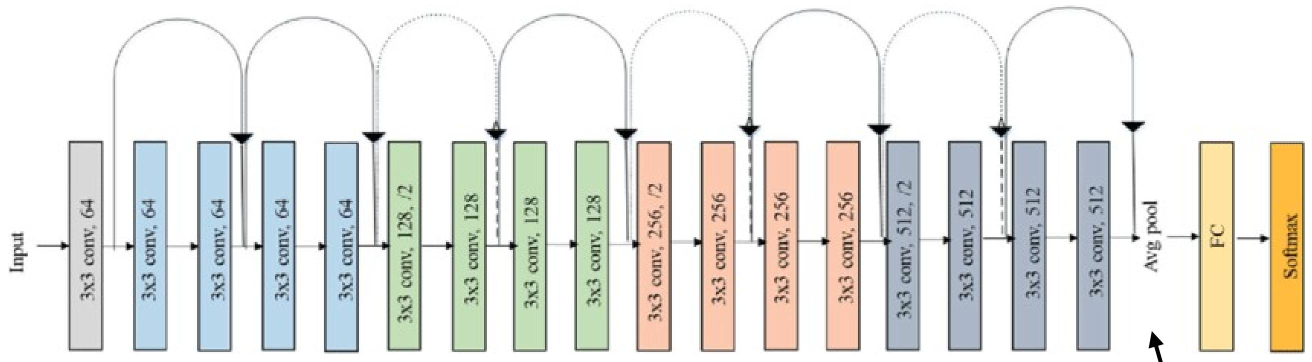
[2]

Continual learning - approaches



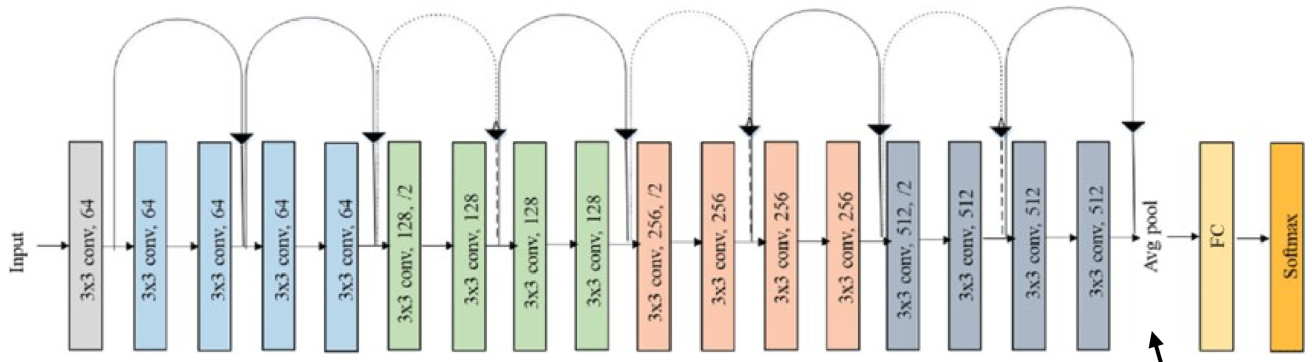
[3]

What happens inside neural network?

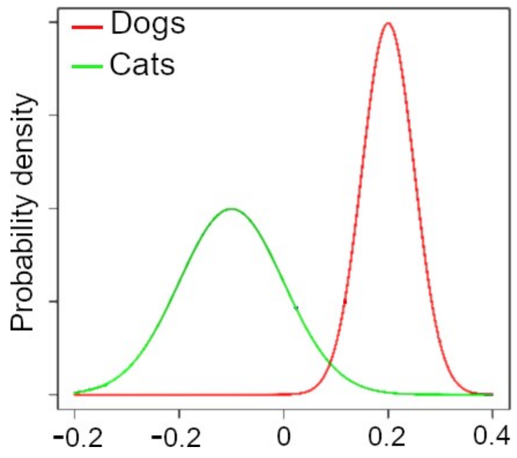


512 dimensional latent space

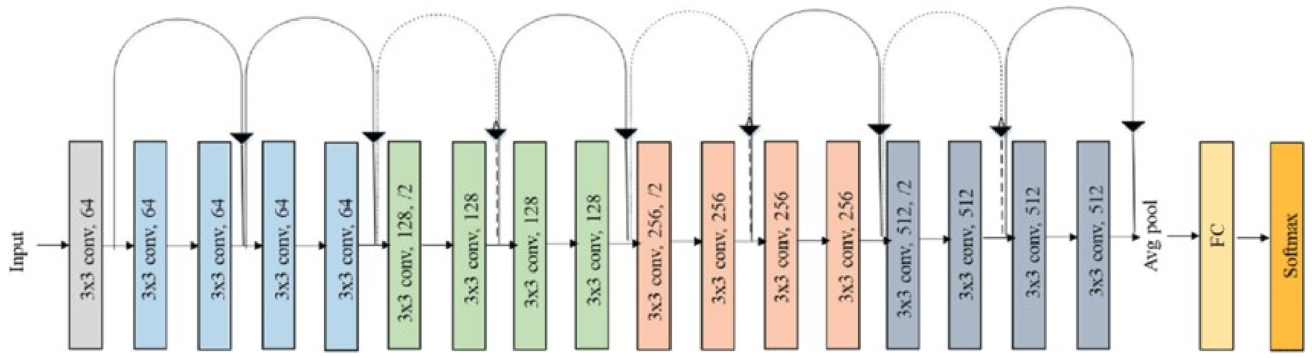
What happens inside neural network?



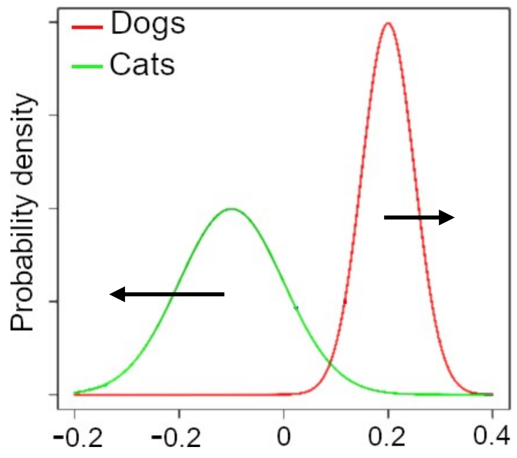
Let's look at the 204th dimension of the latent space



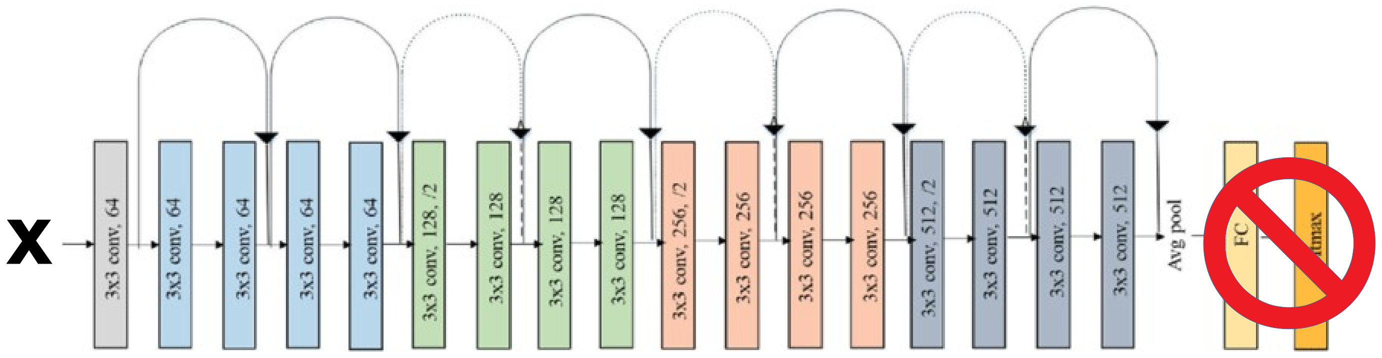
What happens inside neural network?



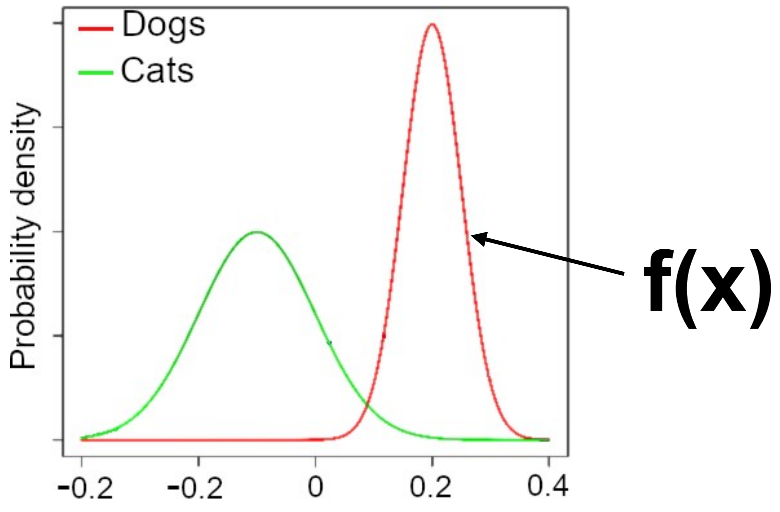
When neural nets forget, these distributions change



Optimal Bayes classifier



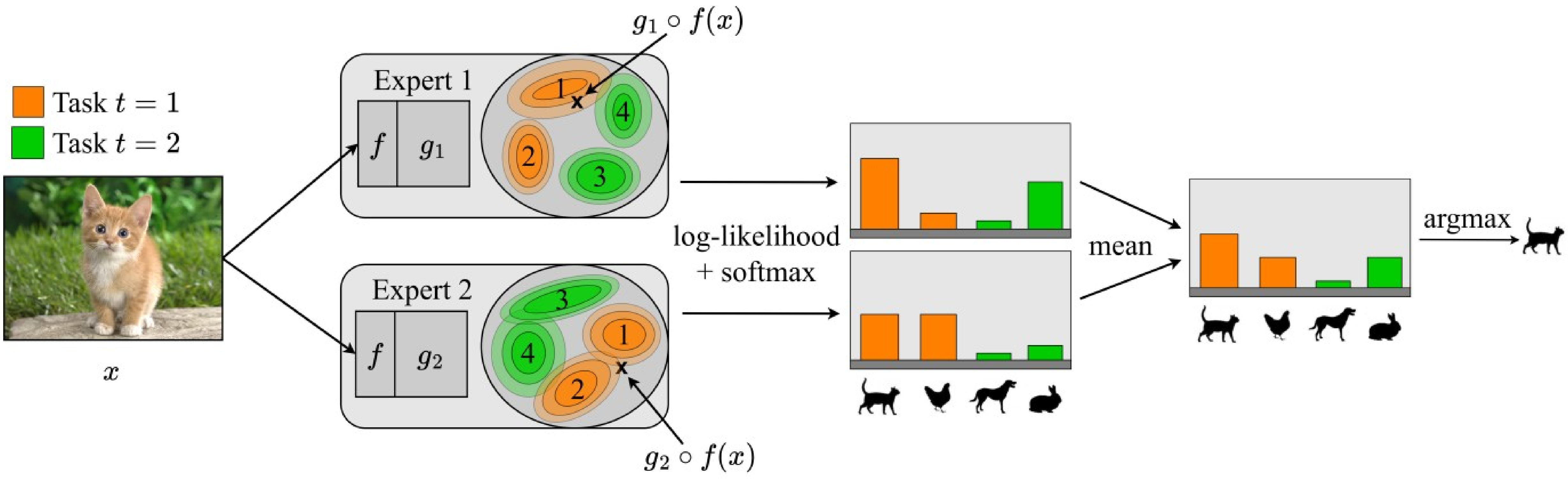
We can replace classification head with Bayes classifier



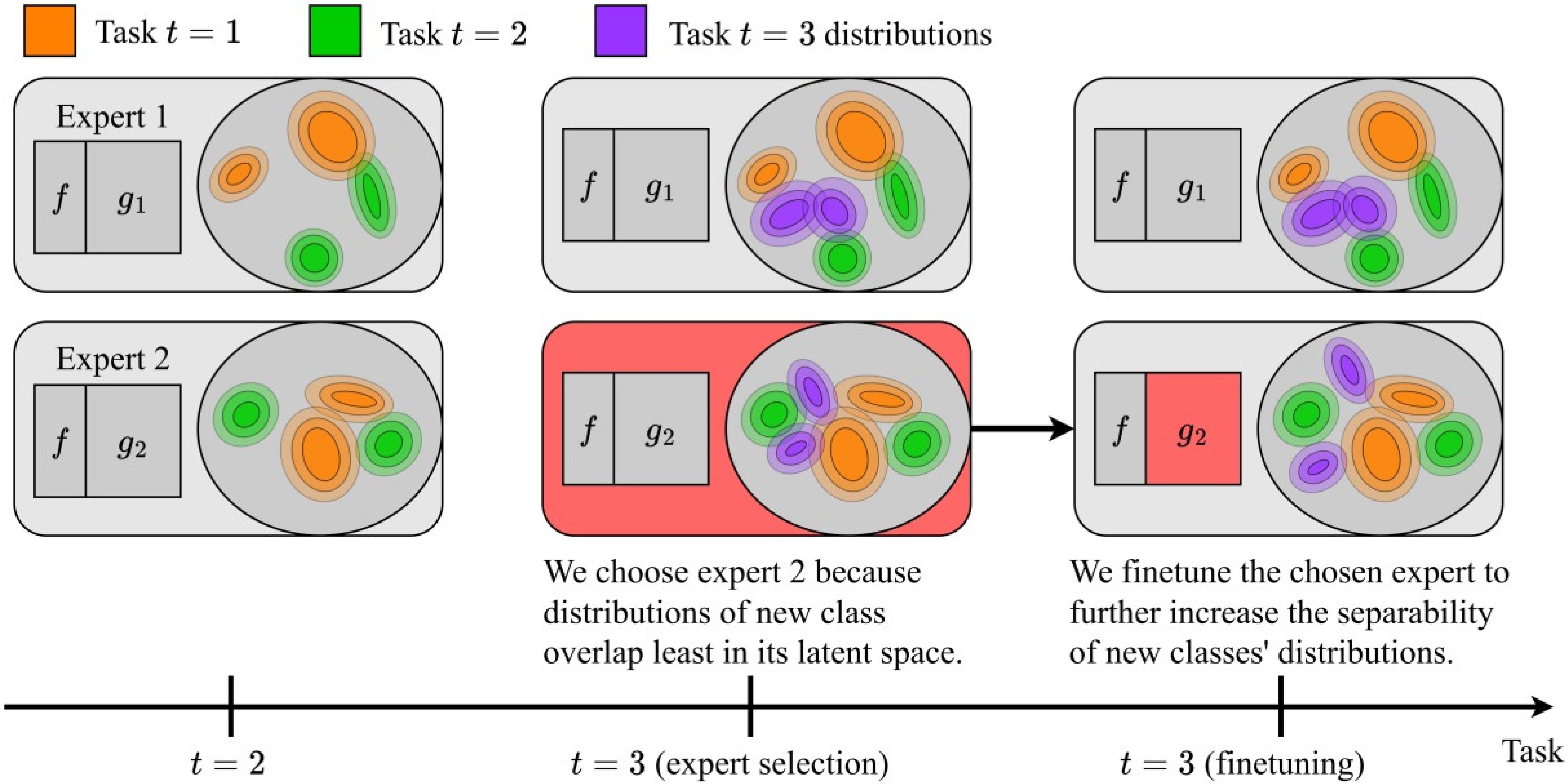
SEED (ICLR'24)

- Ensemble of K experts
- Each expert is a deep network and a set of Multivariate Gaussian distributions representing classes
- For each task select one expert to train its network but create Gausses for all experts
- Diversification is obtained by training experts on different data

SEED - Inference



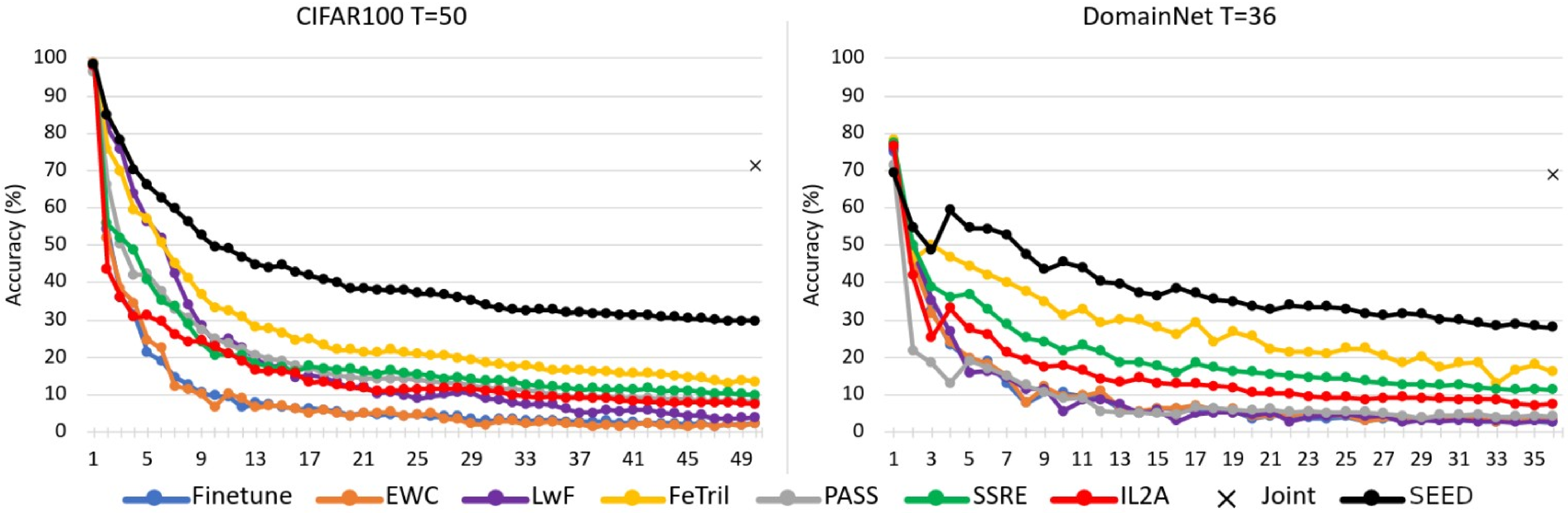
SEED - Training



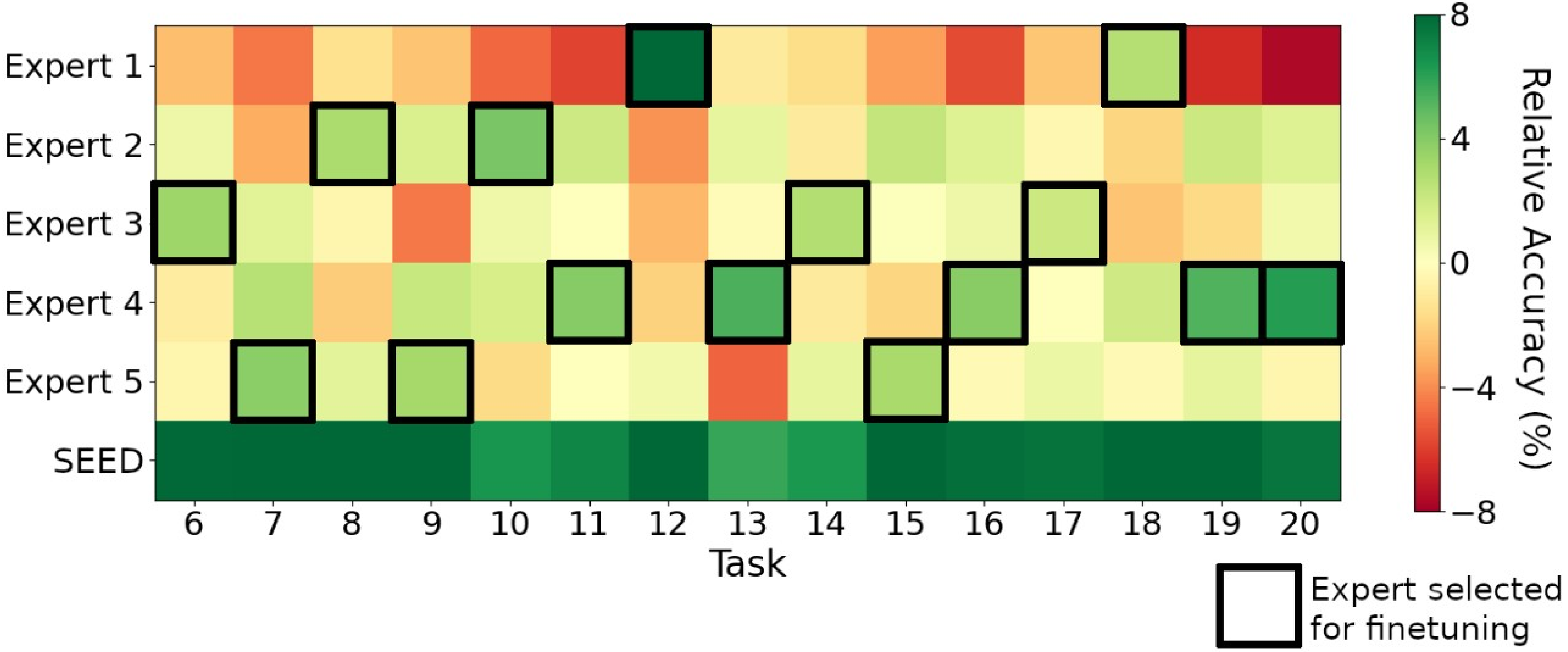
SEED - Results

CIL Method	CIFAR-100 (ResNet32)			DomainNet			ImageNet-Subset
	T=10	T=20	T=50	T=12	T=24	T=36	T=10
Finetune	26.4±0.1	17.1±0.1	9.4±0.1	17.9±0.3	14.8±0.1	10.9±0.2	27.4±0.4
EWC (Kirkpatrick et al., 2017) (PNAS'17)	37.8±0.8	21.0±0.1	9.2±0.5	19.2±0.2	15.7±0.1	11.1±0.3	29.8±0.3
LwF* (Rebuffi et al., 2017) (CVPR'17)	47.0±0.2	38.5±0.2	18.9±1.2	20.9±0.2	15.1±0.6	10.3±0.7	32.3±0.4
PASS (Zhu et al., 2021b) (CVPR'21)	37.8±1.1	24.5±1.0	19.3±1.7	25.9±0.5	23.1±0.5	9.8±0.3	-
IL2A (Zhu et al., 2021a) (NeurIPS'21)	43.5±0.3	28.3±1.7	16.4±0.9	20.7±0.5	18.2±0.4	16.2±0.4	-
SSRE (Zhu et al., 2022) (CVPR'22)	44.2±0.6	32.1±0.9	21.5±1.8	33.2±0.7	24.0±1.0	22.1±0.7	45.0±0.5
FeTrIL (Petit et al., 2023) (WACV'23)	46.3±0.3	38.7±0.3	27.0±1.2	33.5±0.6	33.9±0.5	27.5±0.7	58.7±0.2
SEED	**61.7±0.4**	**56.2±0.3**	**42.6±1.4**	**45.0±0.2**	**44.9±0.2**	**39.2±0.3**	**67.8±0.3**
Joint		71.4±0.3		63.7±0.5	69.3±0.4	69.1±0.1	81.5±0.5

SEED - Results



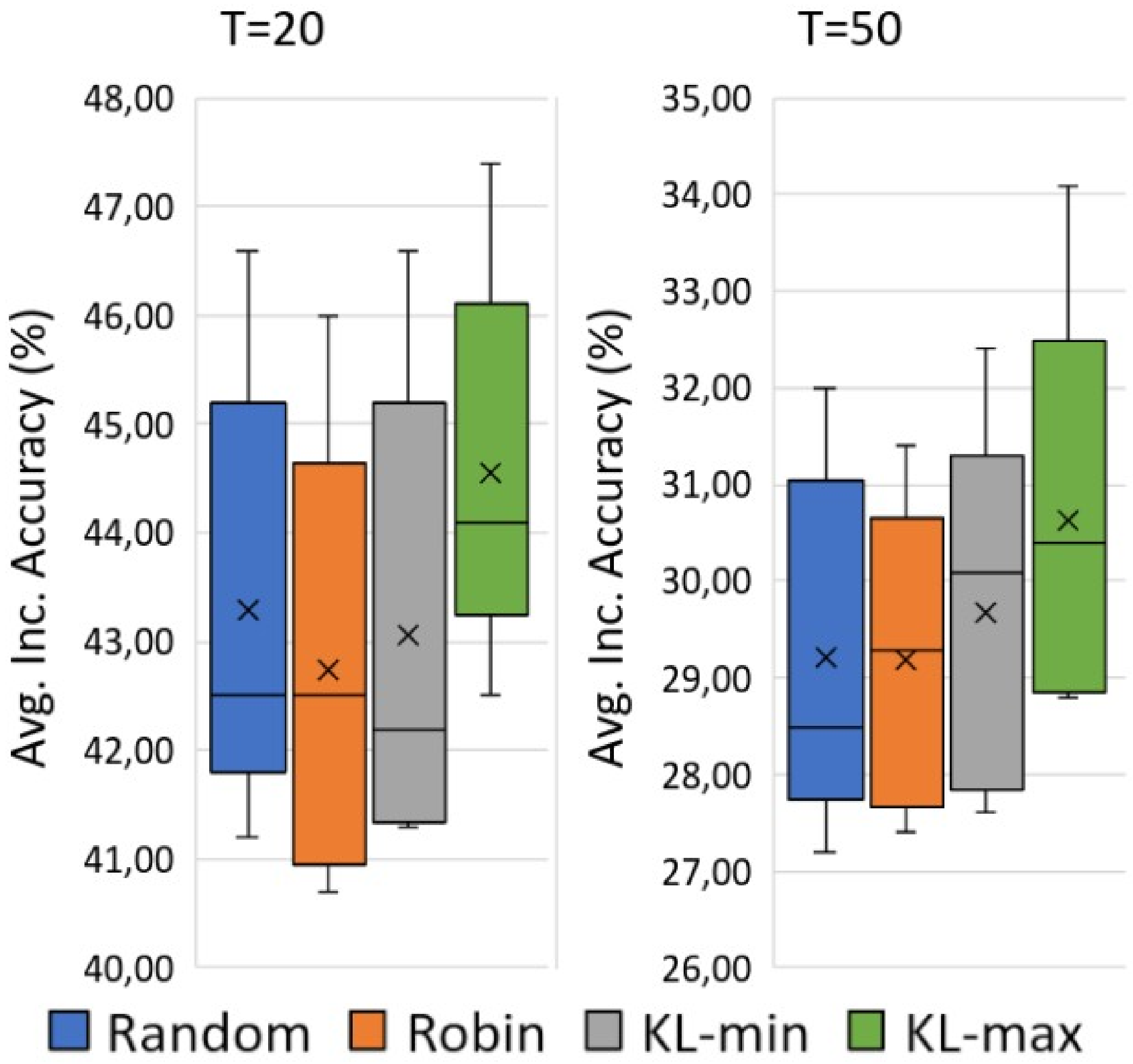
SEED - Diversification of experts



IDEAS NCBR

SEED

How does expert selection strategy affects results?



IDEAS NCBR

Q&A

Link to the paper:

Grzegorz Rypeść
LinkedIn
grzegorz.rypesc@ideas-ncbr.pl

References

[1] https://pl.wikipedia.org/wiki/Hipokamp
[2] van de Ven, Gido M., and Andreas S. Tolias. "Three scenarios for continual learning." arXiv e-prints (2019): arXiv-1904.
[3] Masana, Marc, et al. "Class-incremental learning: survey and performance evaluation on image classification." IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).