Referential Communication in Heterogeneous Communities of Pre-trained Visual Deep Networks

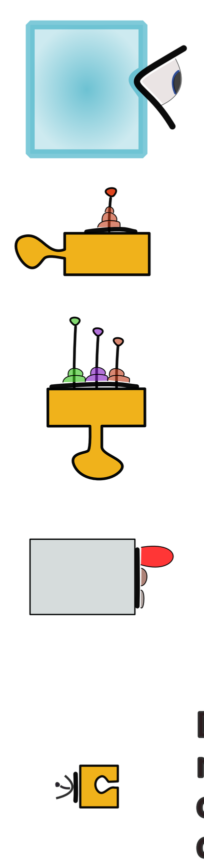
Matéo Mahaut[1], Roberto Dessí[1,2], Francesca Franzon[1], Marco Baroni[1,3]

1-Universitat Pompeu Fabra, 2-Meta AI,3-ICREA

Problem definition and setup

We want different networks to be able to cooperate. Should your smart fridge need to communicate with your new smart microwave from a different brand, can their inner networks work out a way to share information? We investigate how state of the art neural networks might communicate in a group, despite their differences.

Modules



Vision module ***(see Table)*****. Deep Neural Network pretrained on ImageNet1k. Weights are frozen.**

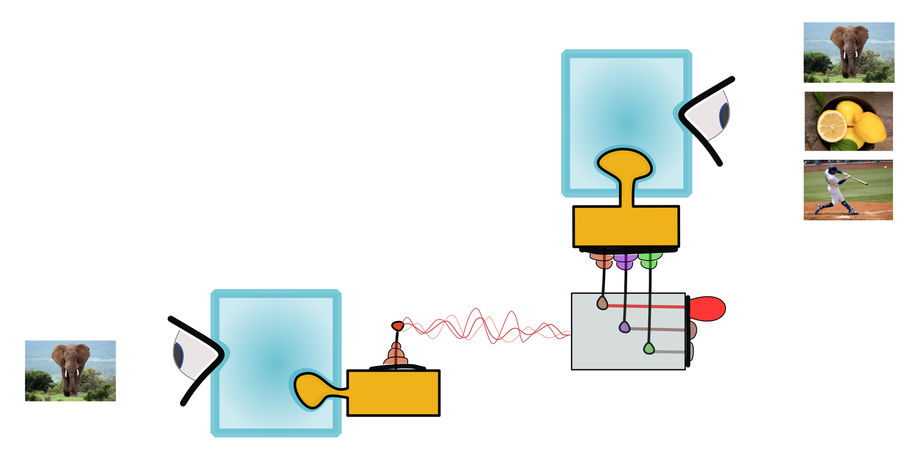
Sender Module. Trainable one layer feed forward neural network.

Mapper module. Trainable two layer feed forward neural network.

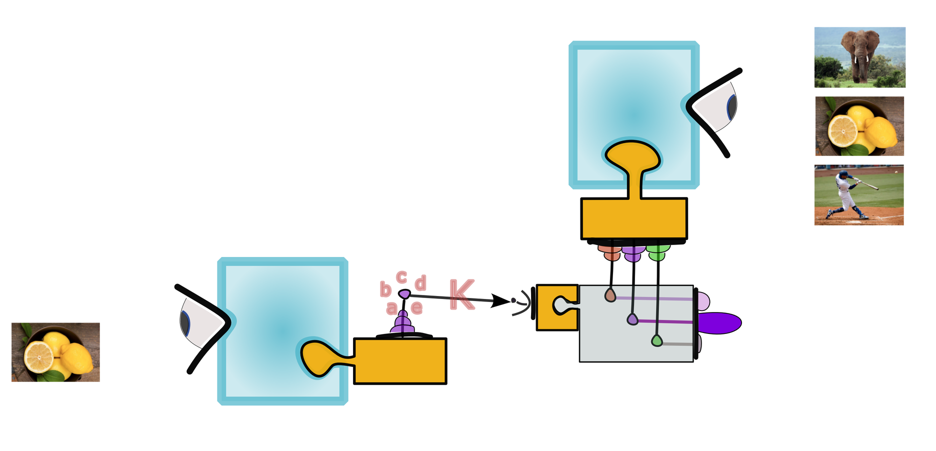
Comparison Module. Applies cosine similarity between inputs received from Mapper and Sender.

Decoder module. Feed forward neural network that maps discrete messages to continuous representations. Only used with discrete communication.

Continuous communication game



Discrete communication game



Datasets

In domain: All training is done on the Imagenet1k Validation set which has not been seen during the vision module's pre-training. 10% of the set is kept for testing.

Out of domain: To test for generalisation capabilities, we select categories from Imagenet 21k for which vision modules never performed the classification task.

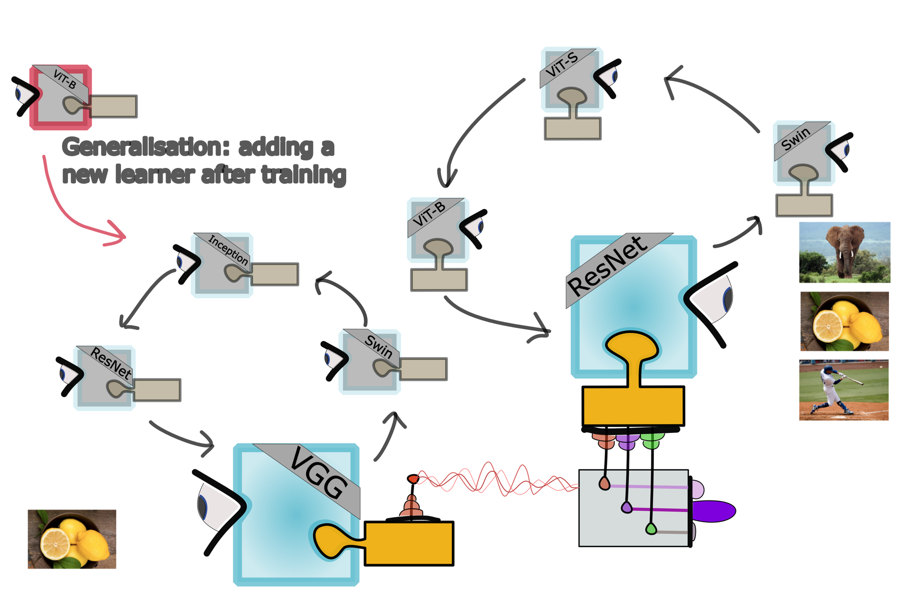
Single class: Using *In domain* images, batches are organised so that vision modules must communicate about images from a single Imagenet class.

Vision modules

Architecture	Type	Training	Parameters
ResNet152	CNN	Supervised	60.2M
Inception	CNN	Supervised	27.2M
VGG 11	CNN	Supervised	132.9M
ViT-B/16	Attention	Supervised	86.6M
ViT-S/16	Attention	Self-supervised	21M
Swin	Attention	Supervised	87.7M

Results

Population communication game



Percentage accuracy on single-class test set

	Discrete	Continuous
Homogeneous	13 ± 4	47 ± 4
Heterogeneous	16 ± 6	37 ± 7
Population	9 ± 2	35 ± 6

Percentage accuracy on OOD test set

	Discrete	Continuous
Homogeneous	43 ± 5	92 ± 5
Heterogeneous	29 ± 7	61 ± 16
Population	26 ± 5	66 ± 15

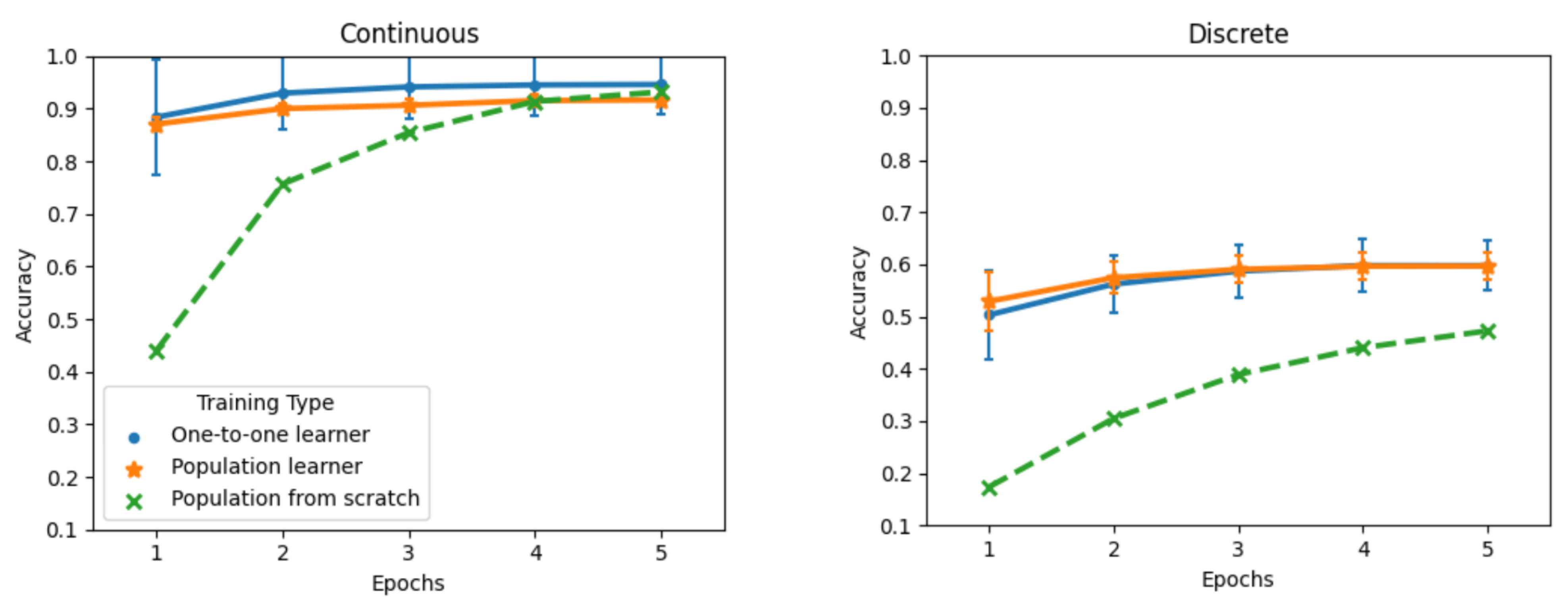
Communication remains possible out of domain, even allowing discrimination at a finer granularity than encoutered during pretraining.

In Domain Communication

	Discrete		Continuous	
	Accuracy	Speed	Accuracy	Speed
Homogeneous	78 ± 0	20 ± 1.5	100 ± 0	3.3 ± 0.94
Heterogeneous	71 ± 4	22 ± 2.3	97 ± 2	3.6 ± 0.62
Population	62 ± 3	23	98 ± 1	27

Image representations can be generalised accross different vision modules to near perfect accuracy, as if they were the same architectures.

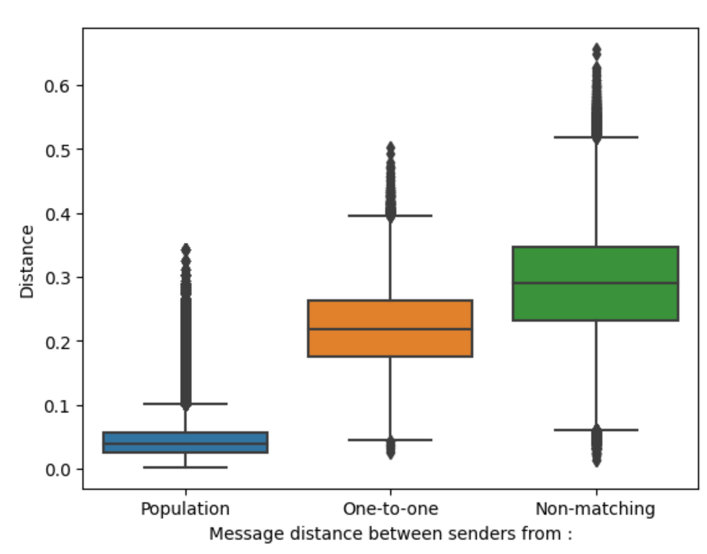
Validation accuracy for agent learning pre-established communication protocol



New agents can easily learn a protocol developped by others. Population training facilitates transfer accross less compatible pairs, resulting in lower variance.

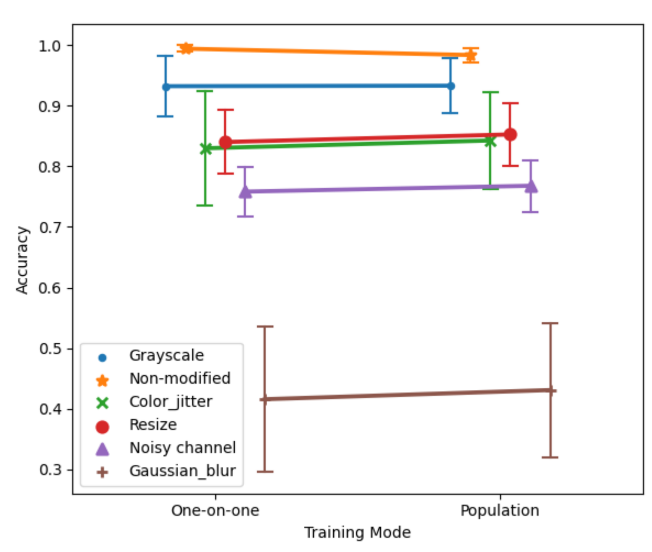
Communication Analysis

Image representation distance



Modules describe the same image (blue, orange) with more similar messages than if they describe different ones (green).

Image modifications



Communication relies on high level image properties, that are stable to classical image transformations.

Take-home messages

Emergent communication allows communication accross architectures, training method, and size despite complex high dimensional data. The trained communication modules:
* Generalise to unseen datasets
* Generalise within a class they did not need to at pretraining
* Can be learnt by new agents

- Continuous communication is easier to implement and performs better, but its gradient reliance makes discrete methods necessary in some use cases.

- Population communication is more stably learnt, to similar accuracies and speeds.