Learning to Compare: Relation Network for Few-Shot Learning

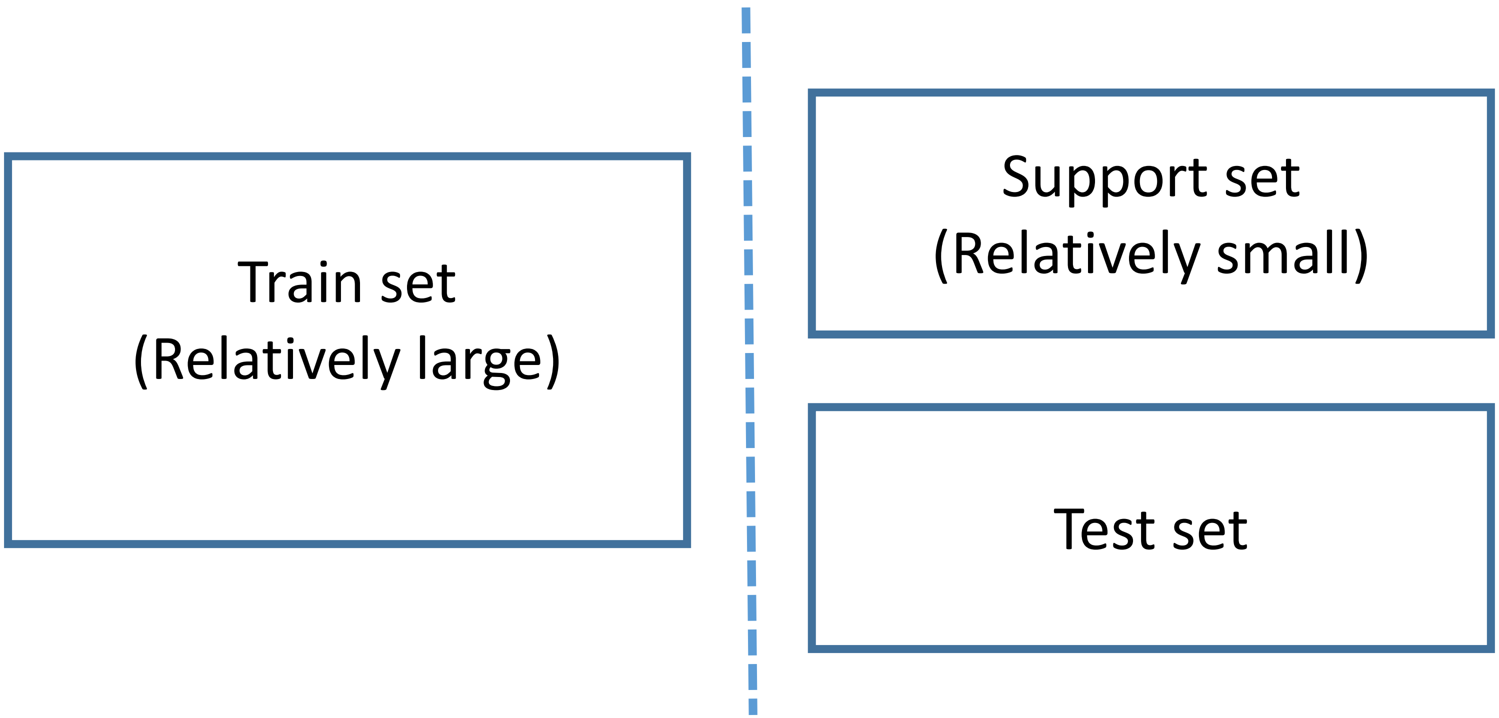
Flood Sung Yongxin Yang[1] Li Zhang[1,2] Tao Xiang[1] Philip H.S. Torr[2] Timothy M. Hospedales[3]
[1]Queen Mary University of London [2]University of Oxford [3]The University of Edinburgh
floodsung@gmail.com {yongxin.yang, david.lizhang, t.xiang}@qmul.ac.uk
philip.torr@eng.ox.ac.uk t.hospedales@ed.ac.uk

Few-shot learning
Meta-learning

https://www.slideshare.net/SimonJohn21/learning-to-compare-relation-network-for-few-shot-learning

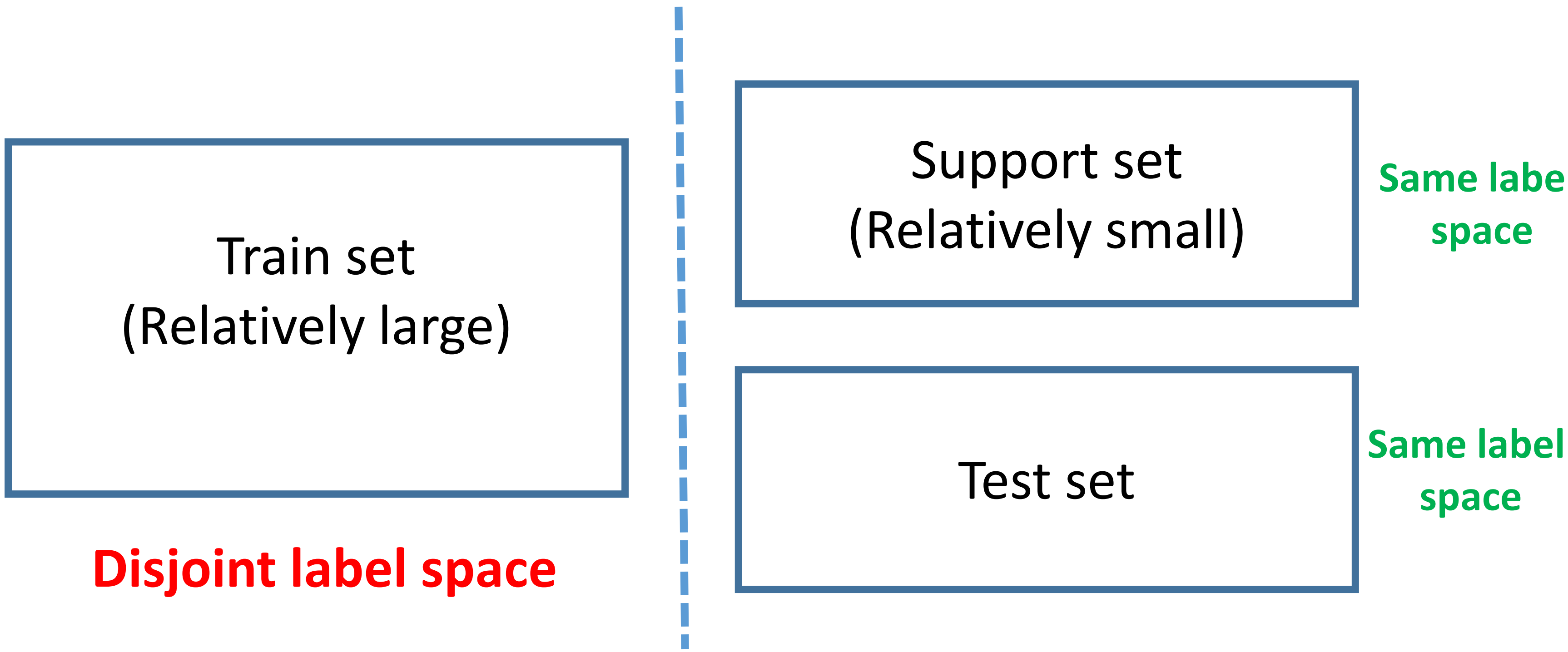
Few-shot learning

Must train a classifier to recognize new classes given few examples from each.



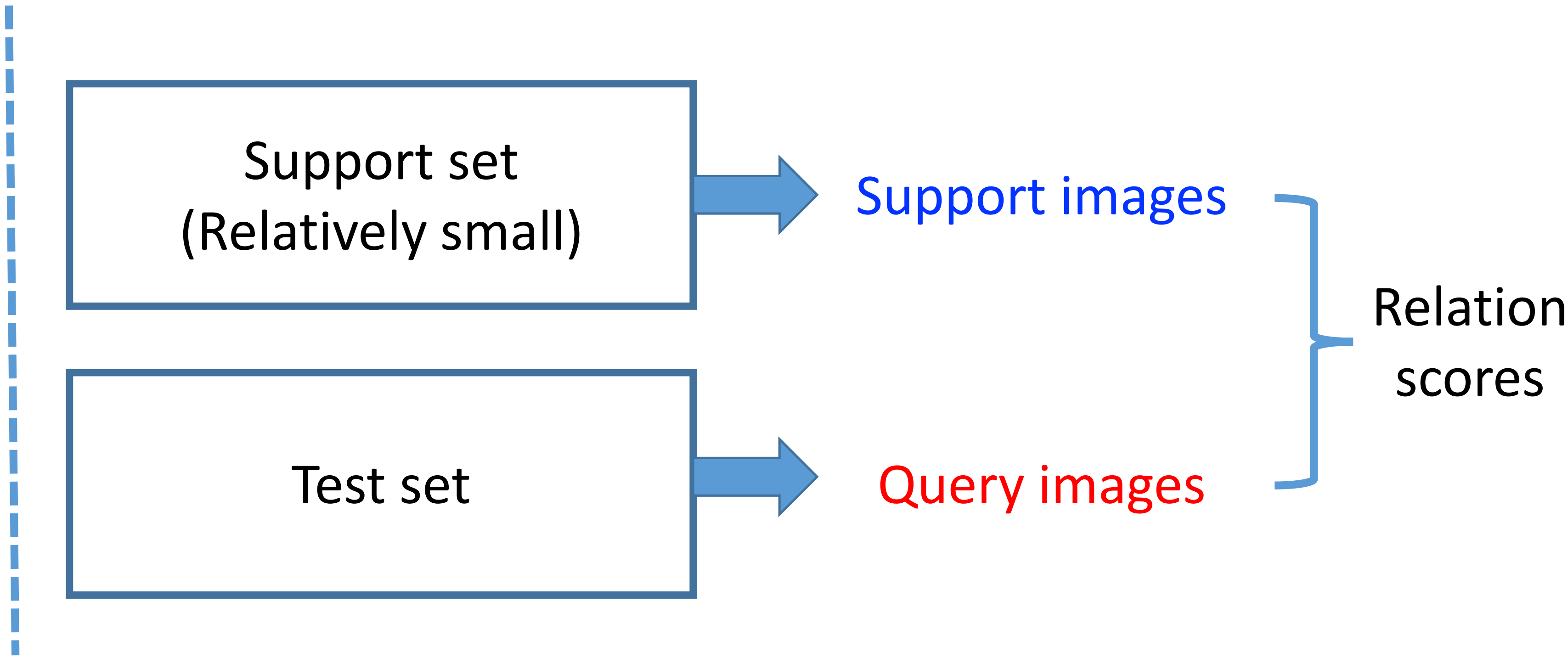
Few-shot learning

Must train a classifier to recognize new classes given few examples from each.



Relation network--classifying by comparison

Relation network (RN) compares support images and query images and makes classification according to the returned "**relation scores**"

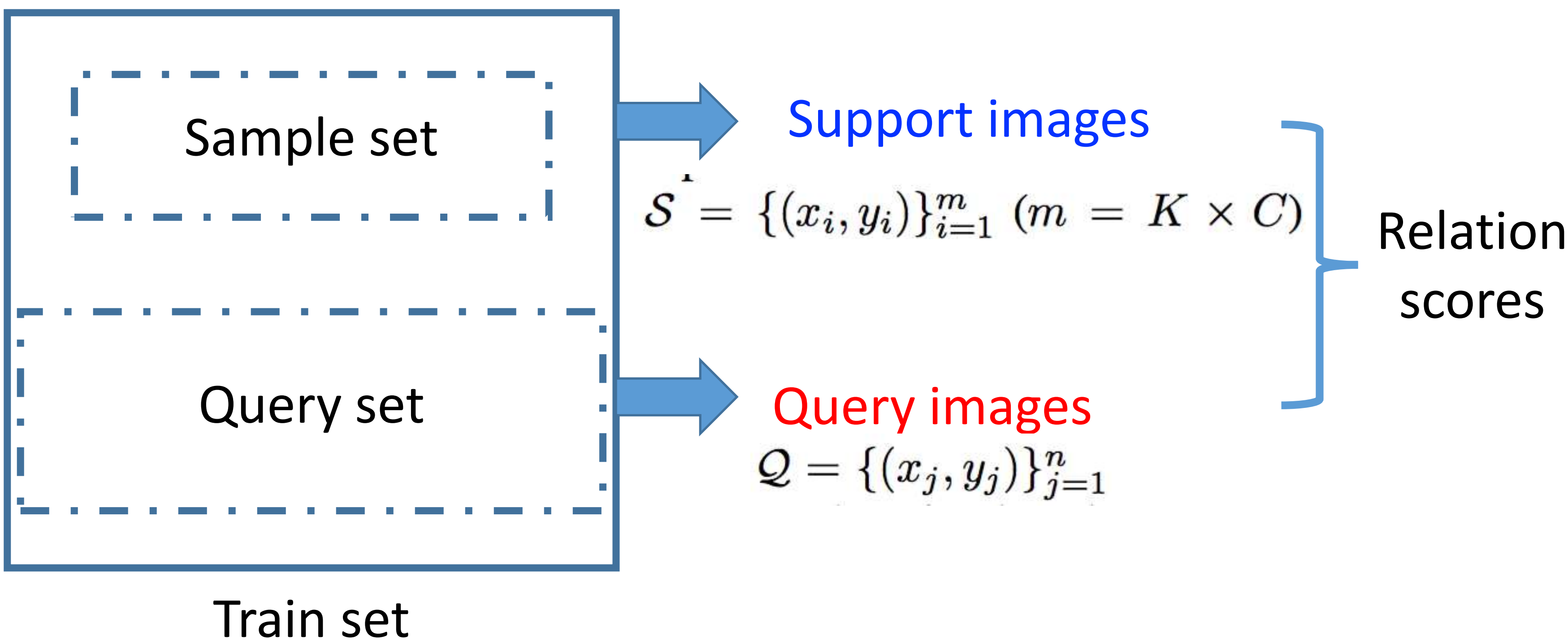


Discussion: Converting classification task into retrieval task?

Relation network--classifying by comparison

Relation network (RN) learn to compare with meta-learning: to mimic the comparison procedure on the training set and learn the model.

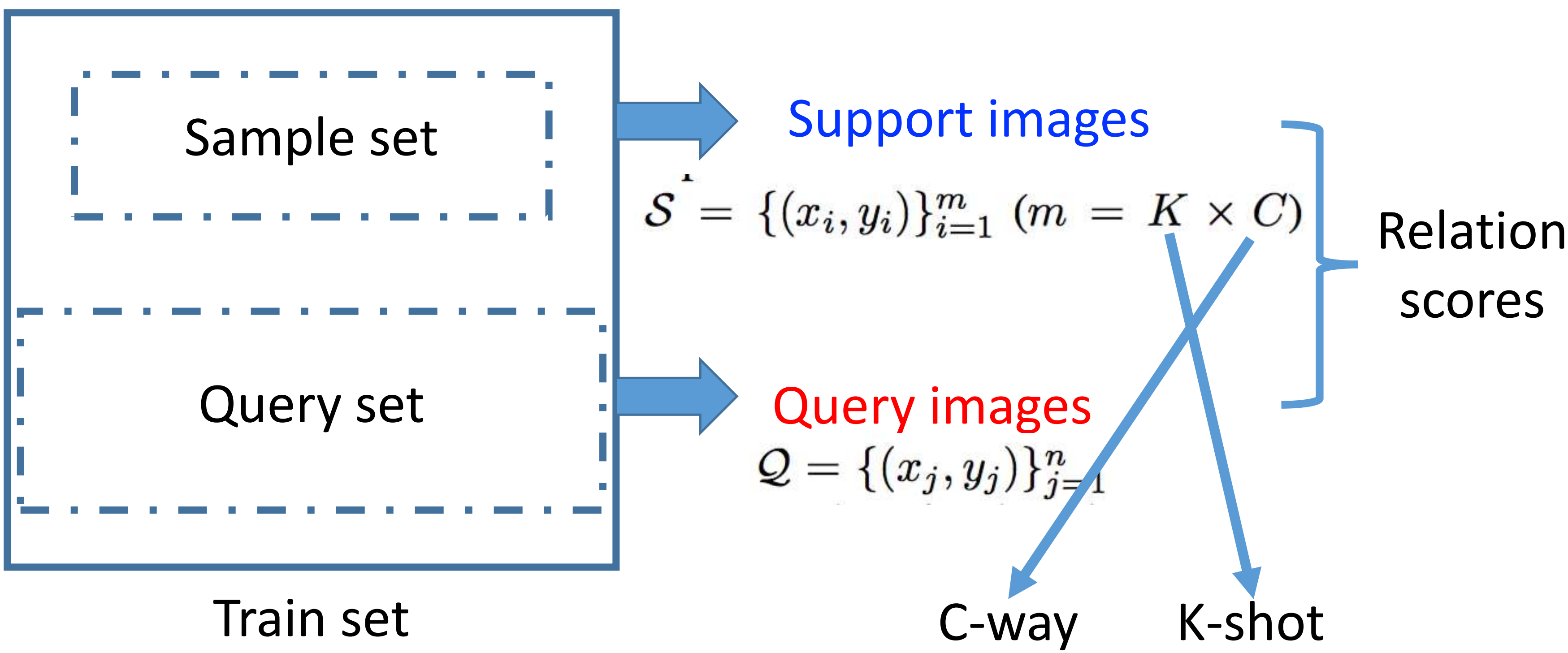
In each episode



Relation network--classifying by comparison

Relation network (RN) learn to compare with meta-learning:
to mimic the comparison procedure on the training set and learn the model.

In each episode



Relation network--structure

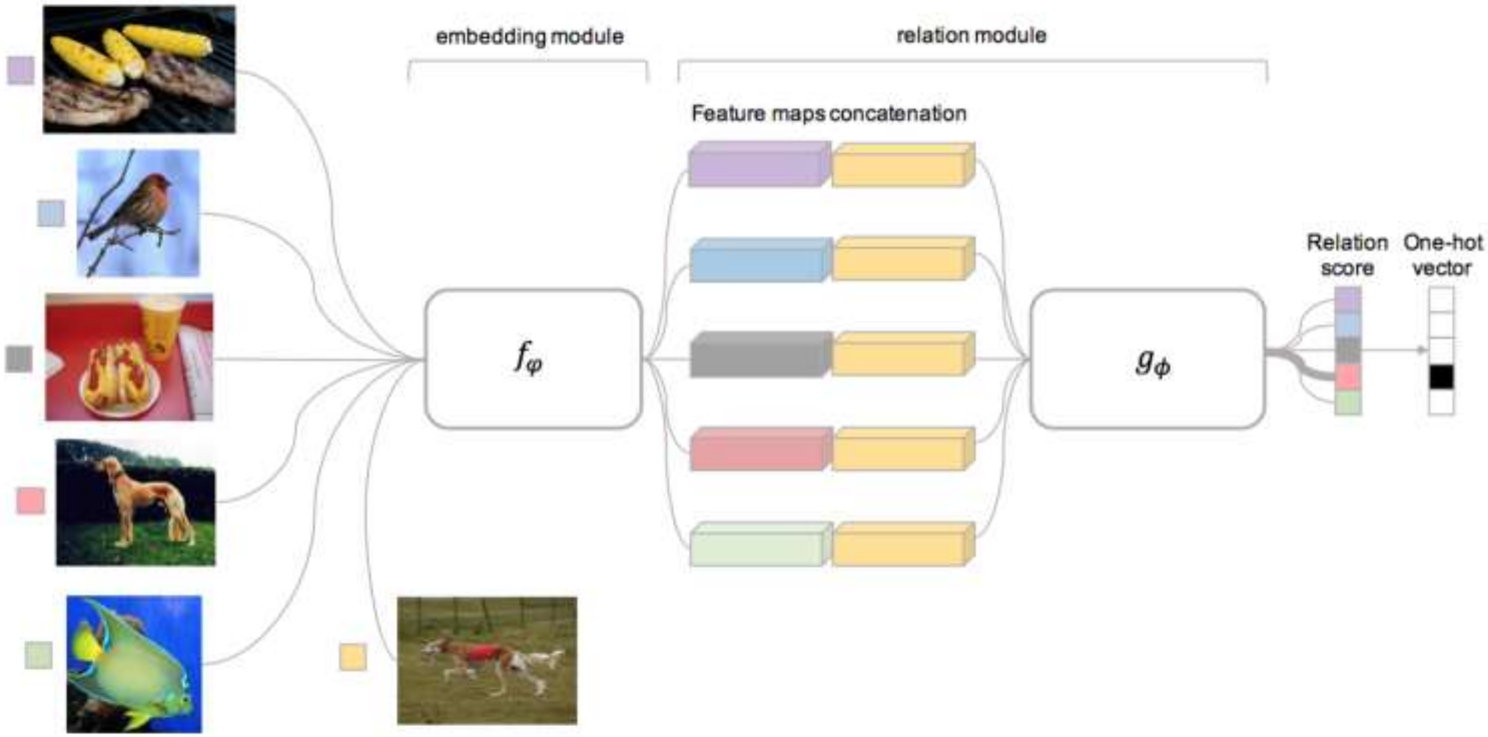
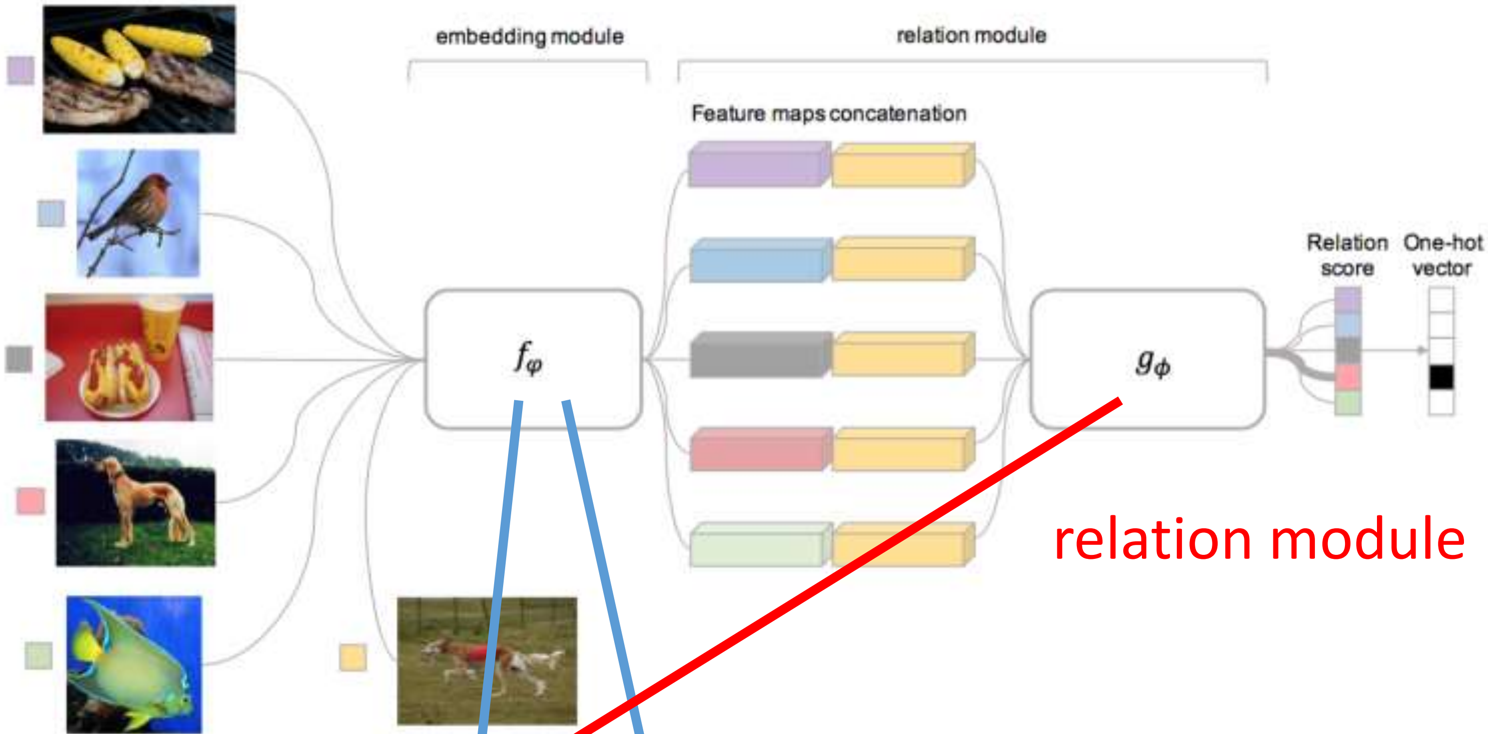


Figure 1: Relation Network architecture with a 5-way 1-shot 1-query example.

$$r_{i,j} = g_\phi(\mathcal{C}(f_\varphi(x_i), f_\varphi(x_j))), \quad i = 1, 2, \ldots, C$$

Relation network--structure



Embedding module

Figure 1: Relation Network architecture with a 5-way 1-shot 1-query example.

$$r_{i,j} = g_\phi(\mathcal{C}(f_\varphi(x_i), f_\varphi(x_j))), \quad i = 1, 2, \ldots, C$$

$r_{i,j}$ is bounded between (0,1) by sigmoid function

Relation network--structure

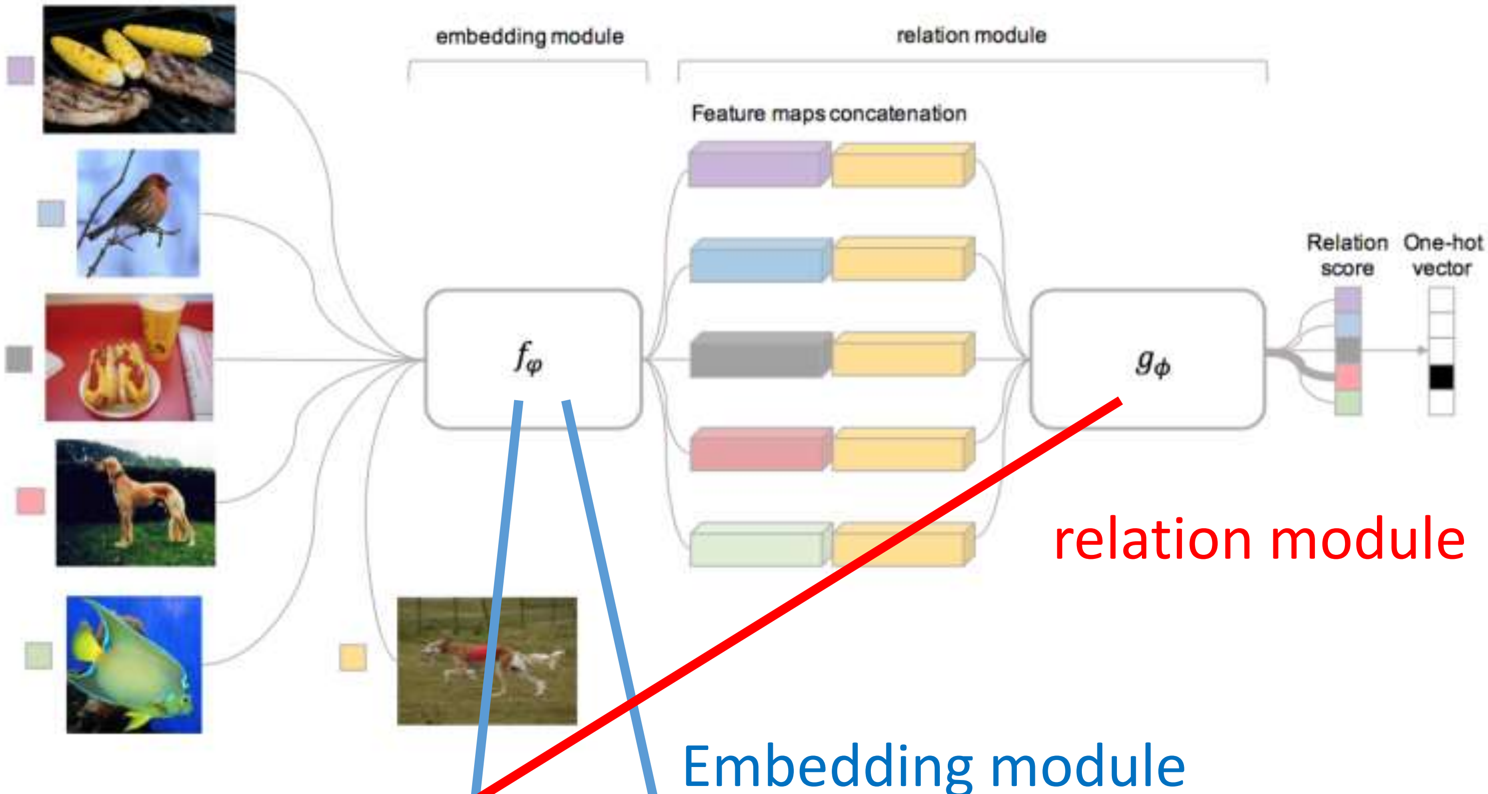


Figure 1: Relation Network architecture with a 5-way 1-shot 1-query example.

$$r_{i,j} = g_\phi(\mathcal{C}(f_\varphi(x_i), f_\varphi(x_j))), \quad i = 1, 2, \ldots, C$$

Optimization target: $$\varphi, \phi \leftarrow \underset{\varphi, \phi}{\operatorname{argmin}} \sum_{i=1}^{m} \sum_{j=1}^{n} (r_{i,j} - \mathbf{1}(y_i == y_j))^2$$

Relation network--structure

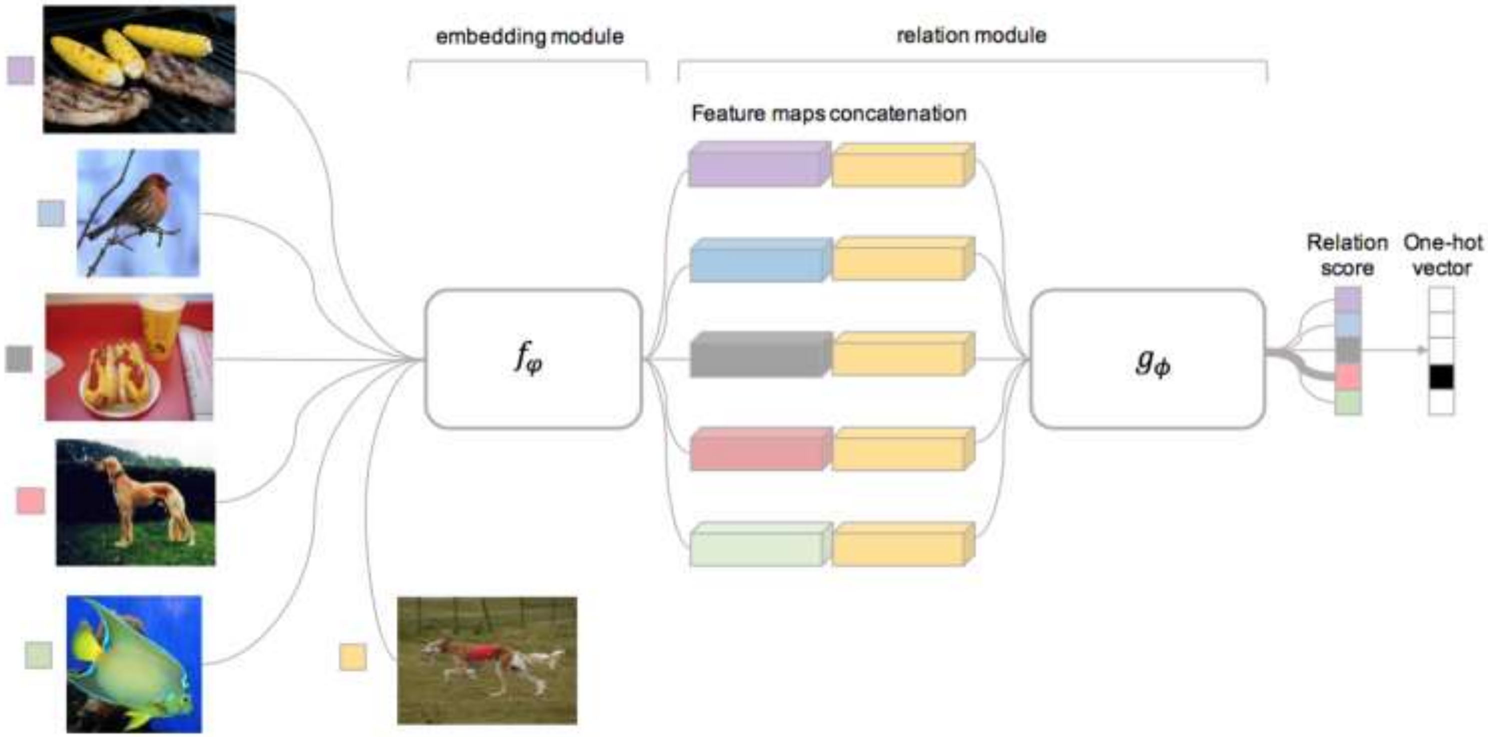
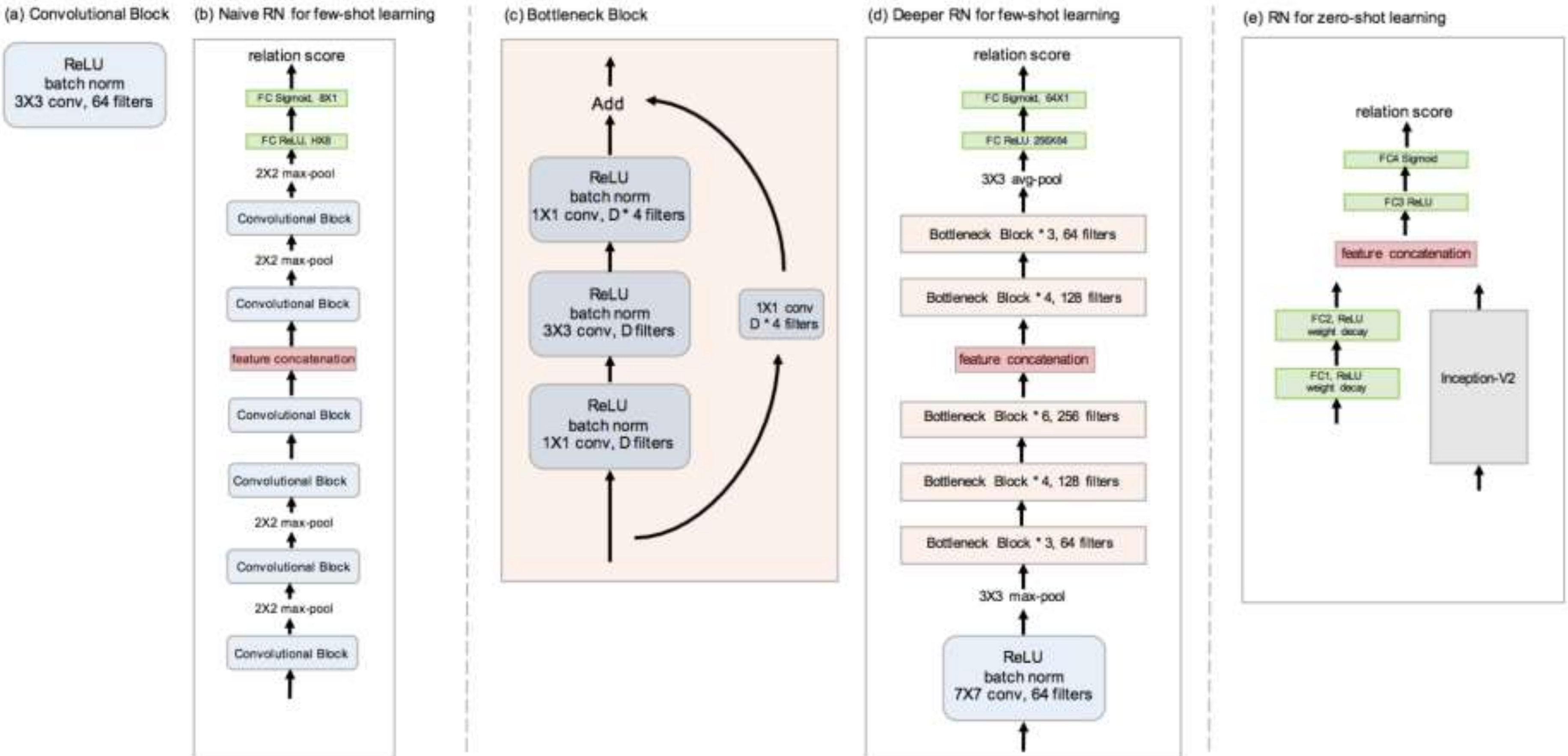


Figure 1: Relation Network architecture with a 5-way 1-shot 1-query example.

A detail for K-shot: K embedded features are pooled by pixel-wise sum operation

Relation network--structure



Detailed structure

Extension to 0-shot learning: different embedding module for sample and query images

Semantic vector images

Experiments

Model	Fine Tune	5-way Acc.		20-way Acc.	
		1-shot	5-shot	1-shot	5-shot
MANN [31]	N	82.8%	94.9%	-	-
CONVOLUTIONAL SIAMESE NETS [18]	N	96.7%	98.4%	88.0%	96.5%
CONVOLUTIONAL SIAMESE NETS [18]	Y	97.3%	98.4%	88.1%	97.0%
MATCHING NETS [38]	N	98.1%	98.9%	93.8%	98.5%
MATCHING NETS [38]	Y	97.9%	98.7%	93.5%	98.7%
SIAMESE NETS WITH MEMORY [16]	N	98.4%	99.6%	95.0%	98.6%
NEURAL STATISTICIAN [8]	N	98.1%	99.5%	93.2%	98.1%
META NETS [26]	N	99.0%	-	97.0%	-
PROTOTYPICAL NETS [35]	N	98.8%	99.7%	96.0%	98.9%
MAML [10]	Y	98.7 ± 0.4%	**99.9 ± 0.1%**	95.8 ± 0.3%	98.9 ± 0.2%
RELATION NET	N	**99.6 ± 0.2%**	**99.8± 0.1%**	**97.6 ± 0.2%**	**99.1± 0.1%**

Table 1: Omniglot few-shot classification. Results are accuracies averaged over 1000 test episodes and with 95% confidence intervals where reported. The best-performing method is highlighted, along with others whose confidence intervals overlap. '-': not reported.

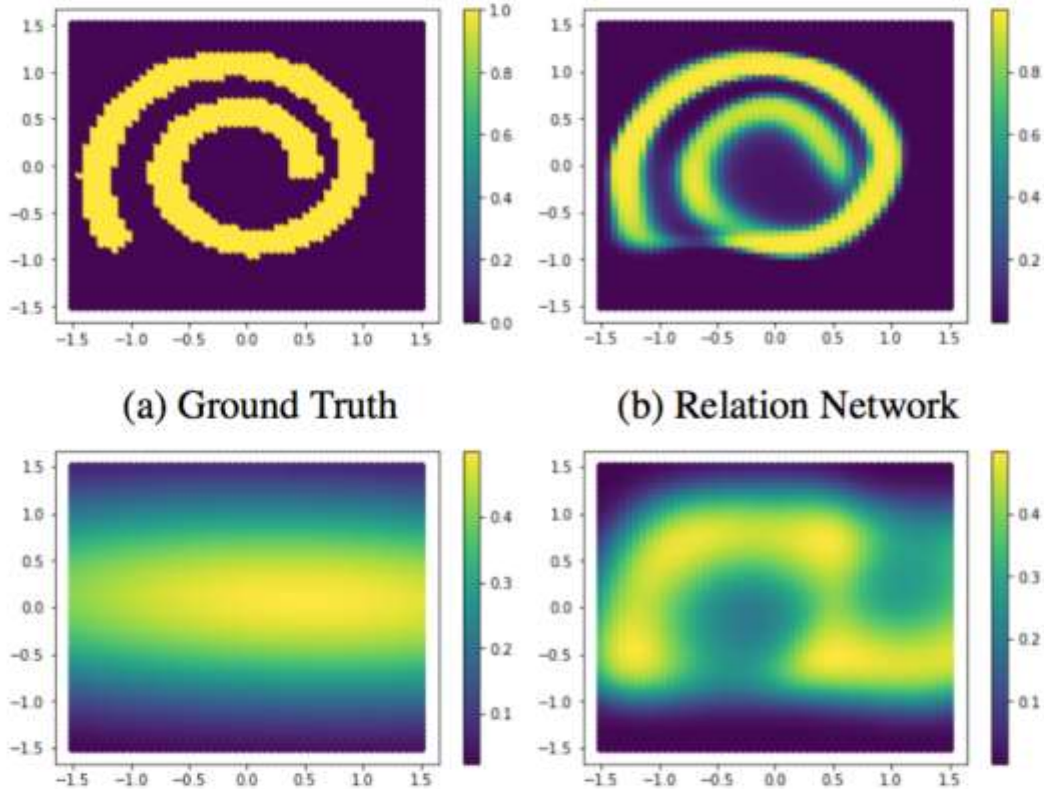
Why does RN work?

Both deep feature embedding and deep distance metric are learnable.

The concatenation operation is in relatively bottom layer

When training a Siamese network or a triplet network, we apply metric constraint on a specified feature and then use the Euclidean distance (or other fixed metric) for metric for inference.

Why does RN work?



(a) Ground Truth (b) Relation Network

(c) Metric Learning (d) Metric + Embedding

Figure 3: An example relation learnable by Relation Network and not by non-linear embedding + metric learning.

2D data space

Why does RN work?

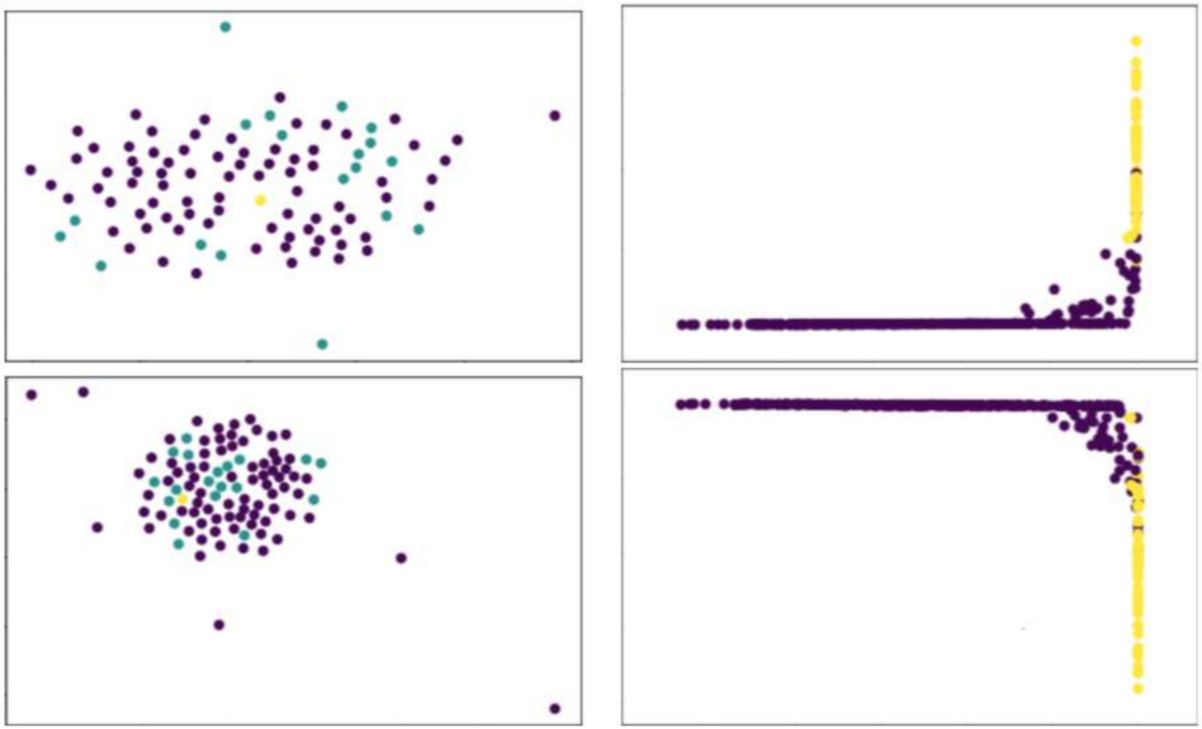


Figure 4: Example Omniglot few-shot problem visualisations. Left: Matched (cyan) and mismatched (magenta) sample embeddings for a given query (yellow) are not straightforward to differentiate. Right: Matched (yellow) and mismatched (magenta) relation module pair representations are linearly separable.

The feature embeddings are difficult to separate.

The relation module pair representations are linearly separable

Why does RN work? My guess

1) Feature concatenation operation in very early stage (bottom layers)

2) K sample images to mimic the K-shot

3) Converting classification to ”comparison”, which is a semi-parameter model approach.