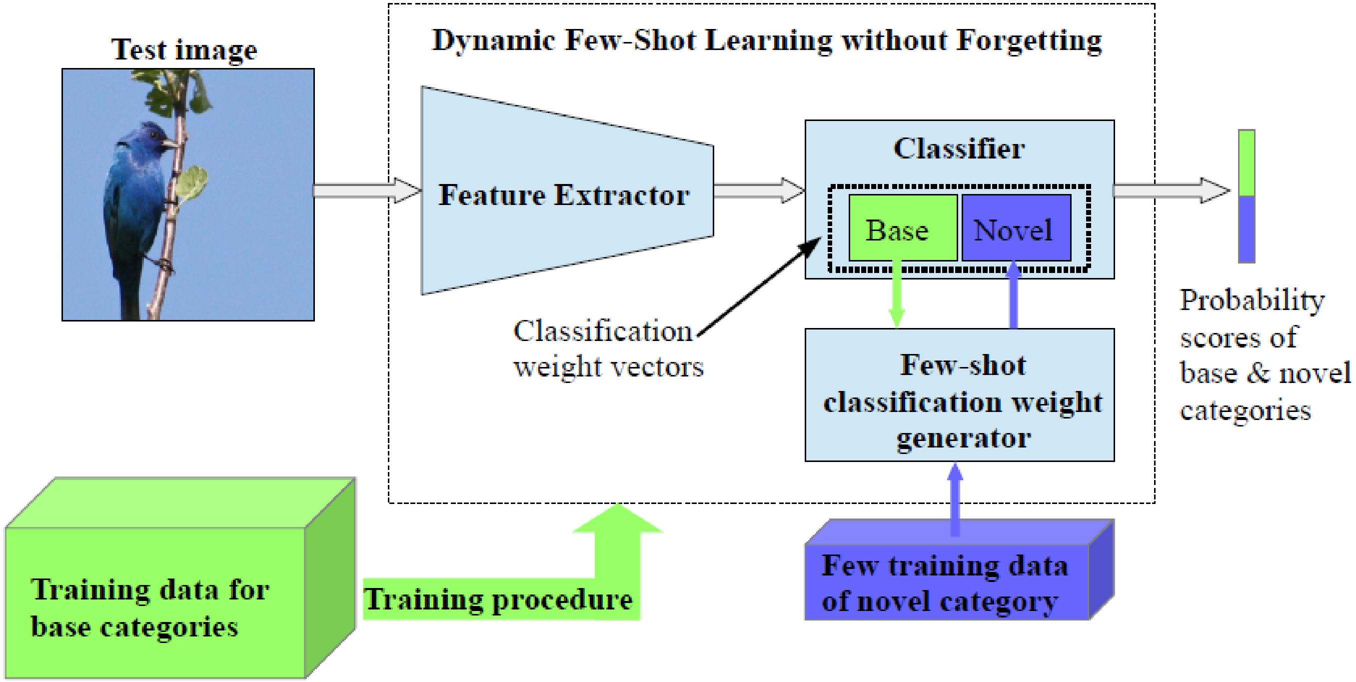
Dynamic Memory Induction Network for Few-Shot Text Classification

本篇论文由北京阿里巴巴达摩院与加拿大皇后大学联合发表于2020ACL

Content

1.Background



1. Few-shot classification-weight generator based on attention
2. Cosine-similarity based ConvNet recognition model.

Spyros Gidaris and Nikos Komodakis. 2018. Dynamic few-shot visual learning without forgetting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4367–4375

1.1 Key Challenges

Problems

1. data sparseness
2. key information is lost to induce class-level representation
3. static memory
4. instance-level diversity

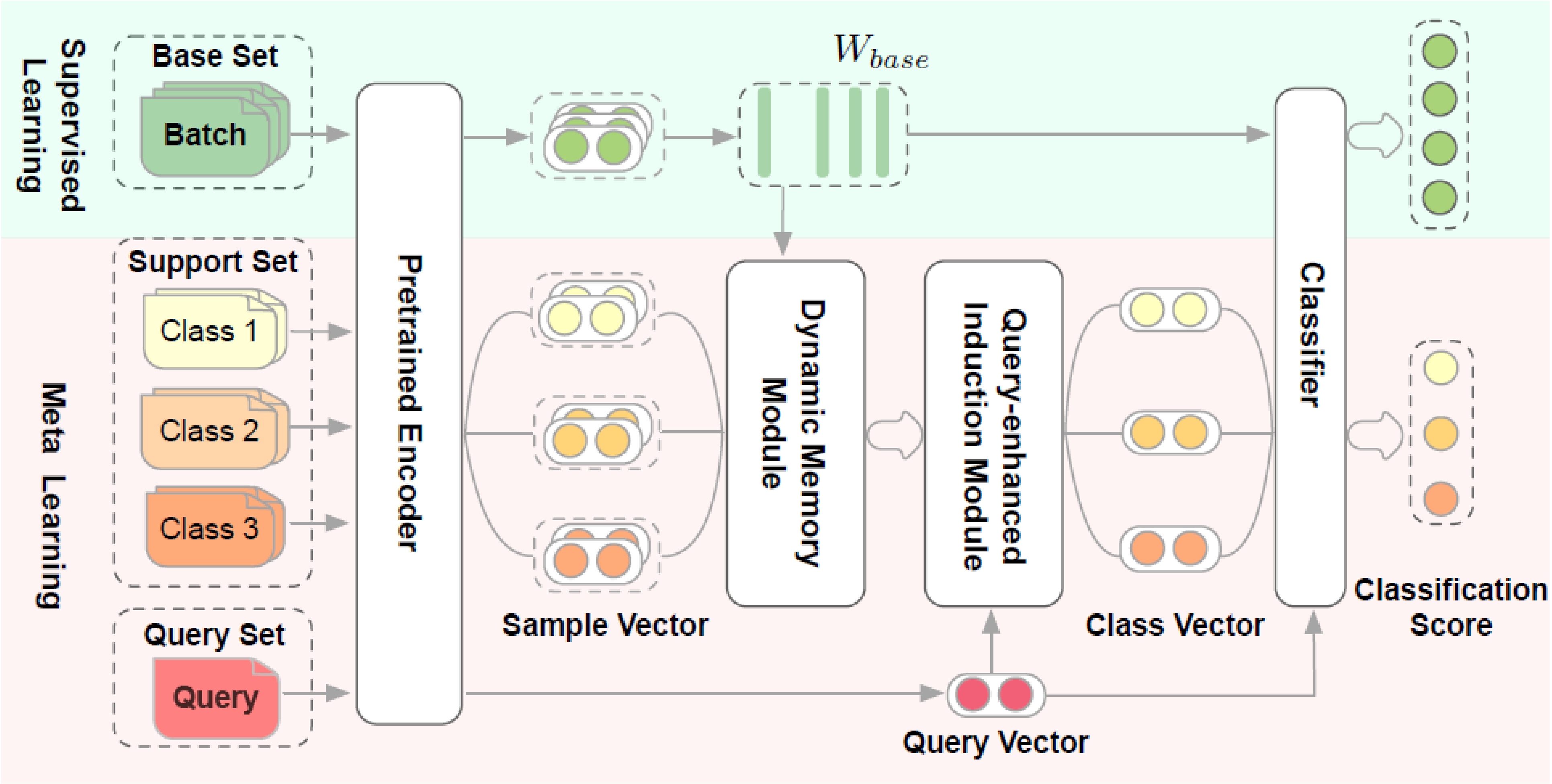
Solutions

1. data augmentation
2. extract transferable knowledge by meta-learning
3. memory component
4. query-aware methods

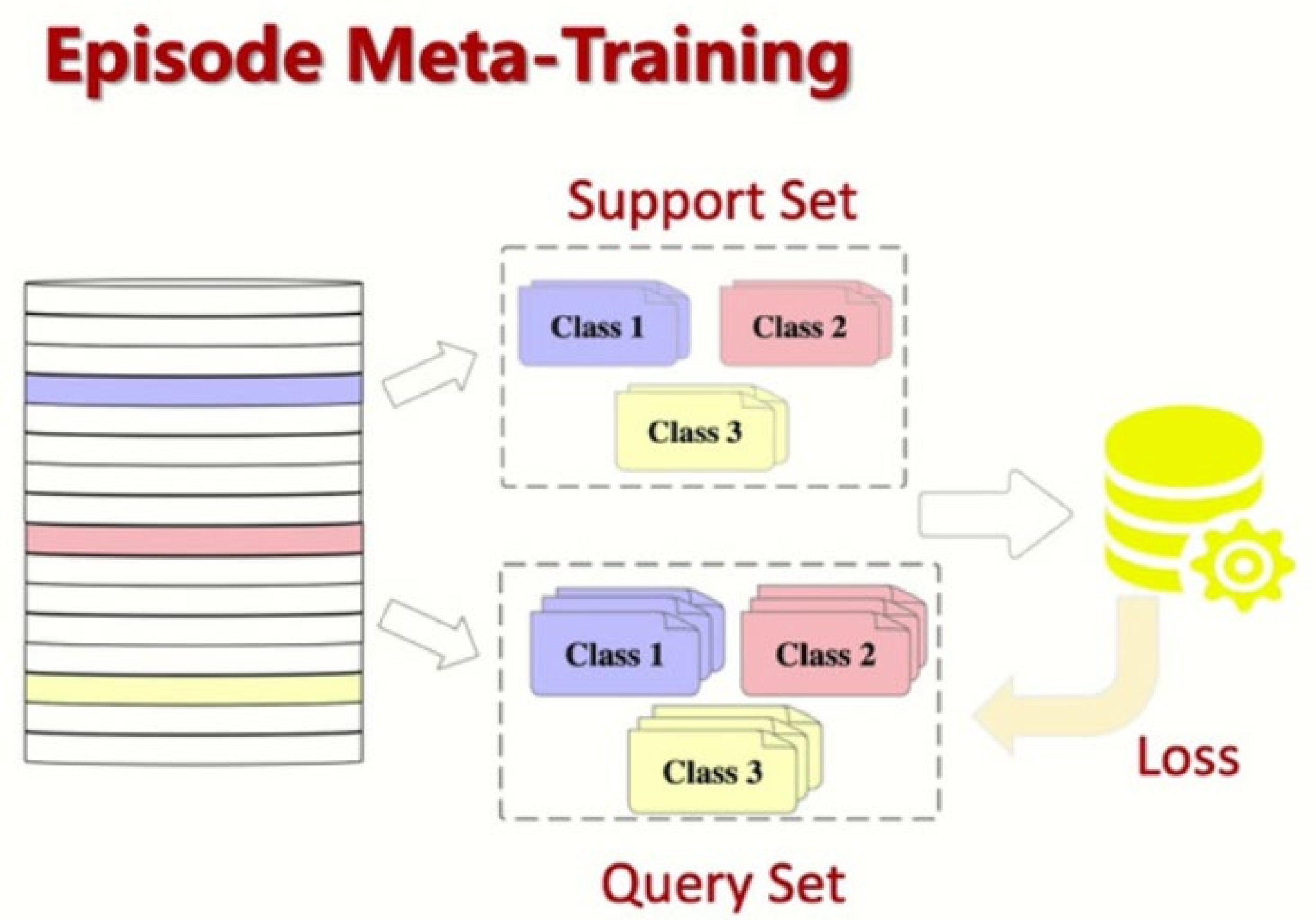
1. J. Salamon and J. P. Bello, Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification, in IEEE Signal Processing Letters
2. Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In AAAI-19

2.Dynamic Memory Induction Networks

DMIN

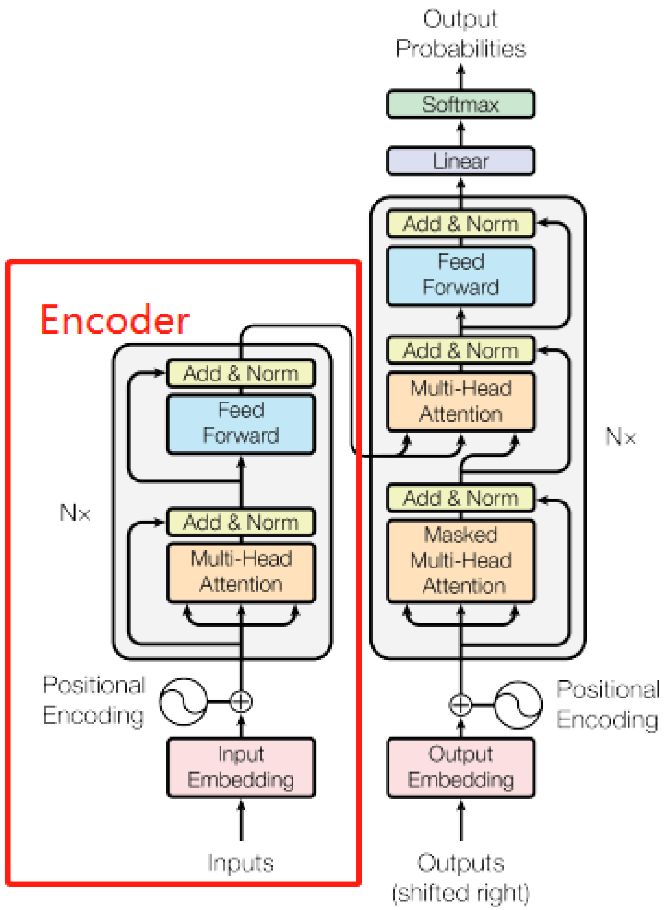


2.1 Meta-Training Episode



- Select C classes
- C-way K-Shot Support Set
- C-way Query set

2.2 Pre-trained Encoder



BERT: a multi-layer bidirectional Transformer encoder based on the original Transformer model

Using Google BERT-Base model as text encoder, fine-tune the model in training procedure

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding https://arxiv.org/abs/1810.04805

Attention Is All You Need https://arxiv.org/abs/1706.03762

2.3 Dynamic Memory Routing Process

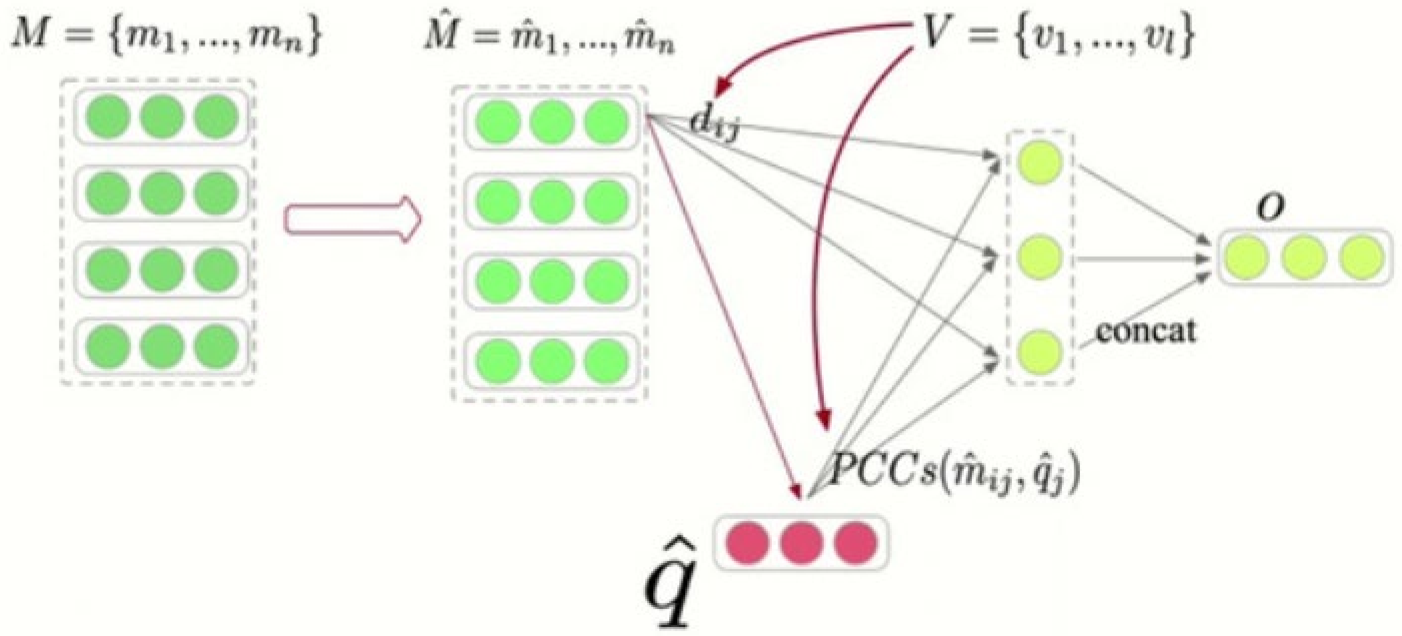
DMR

Algorithm 1 Dynamic Memory Routing Process

Require: r, q and memory $M = \{m_1, m_2, ..., m_n\}$

Ensure: $v = v_1, v_2, ..., v_l$, q'

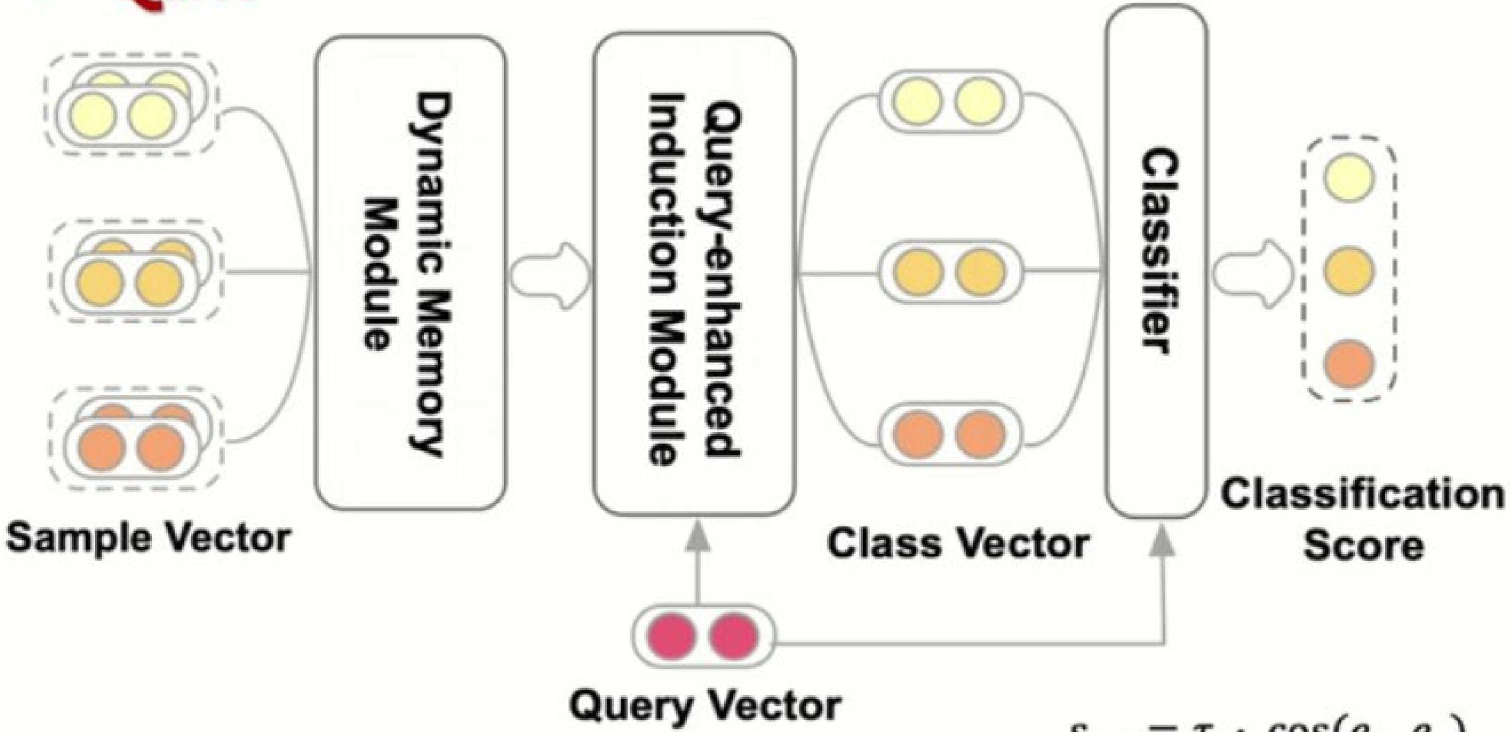
```
for all m_i, v_j do
    m̂_ij = squash(W_j m_i + b_j)
    q̂_j = sqush(W_j q + b_j)
    α_ij = 0
    p_ij = tanh(PCCs(m̂_ij, q̂_j))
end for
for r iterations do
    d_i = softmax(α_i)
    v̂_j = Σ_{i=1}^{n} (d_ij + p_ij) m̂_ij
    v_j = squash(v̂_j)
    for all i, j: α_ij = α_{i,j} + p_ij m̂_ij v_j
    for all j: q̂_j = (q̂_j + v_j) / 2
    for all i, j: p_ij = tanh(PCCs(m̂_ij, q̂_j))
end for
q' = concat[v]
Return q'
```



$$e'_{c,s} = DMR(W_{base}, e_{c,s})$$

2.4 Query-enhanced Induction Module / Similarity Classifier

DMM -> QIM



$$s_{q,c} = \tau \cdot \cos(e_q, e_c)$$

$$e'_{c,s} = DMR(W_{base}, e_{c,s}) \qquad e_c = DMR(\{e'_{c,s}\}_{s=1,\ldots,K}, e_q)$$

3. Experiments and Analysis

Model	5-way Acc.		10-way Acc.	
	1-shot	5-shot	1-shot	5-shot
BERT	30.79±0.68	63.31±0.73	23.48±0.53	61.18±0.82
ATAML	54.05±0.14	72.79±0.27	39.48±0.23	61.74±0.36
Rel. Net	59.19±0.12	78.35±0.27	44.69±0.19	67.49±0.23
Ind. Net	60.97±0.16	80.91±0.19	46.15±0.26	69.42±0.34
HATT	60.40±0.17	79.46±0.32	47.09±0.28	68.58±0.37
LwoF	63.35±0.26	78.83±0.38	48.61±0.21	69.57±0.35
DMIN	**65.72**±0.28	**82.39**±0.24	**49.54**±0.31	**72.52**±0.25

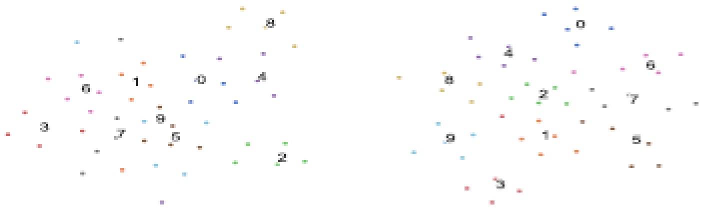
Table 1: Comparison of accuracy (%) on miniRCV1 with standard deviations.

Model	5-way Acc.		10-way Acc.	
	1-shot	5-shot	1-shot	5-shot
BERT	38.06±0.27	64.24±0.36	29.24±0.19	64.53±0.35
ATAML	79.60±0.42	88.53±0.57	63.52±0.34	77.36±0.57
Rel. Net	79.41±0.42	87.93±0.31	64.36±0.58	78.62±0.54
Ind. Net	81.28±0.26	89.67±0.28	64.53±0.38	80.48±0.25
HATT	81.57±0.47	89.27±0.58	65.75±0.61	81.53±0.56
LwoF	79.52±0.29	87.34±0.34	65.04±0.43	80.69±0.37
DMIN	**83.46**±0.36	**91.75**±0.23	**67.31**±0.25	**82.84**±0.38

Table 2: Comparison of accuracy(%) on ODIC with standard deviations.

Model	Iteration	1 Shot	5 Shot
w/o DMM	3	81.79	90.19
w/o QIM	3	82.37	90.57
DMIN	1	82.70	90.92
DMIN	2	82.95	91.18
DMIN	3	**83.46**	**91.75**

Table 3: Ablation study of accuracy (%) on ODIC in a 5-way setup.



(a) Before DMM (b) After DMM

Figure 2: Effect of the Dynamic Memory Module in a 10-way 5-shot setup.

4. Personal Conclusions

- find a mistake in this paper
- the DMIN can also be used in VC
- the DMR of DMIN solves the problem of static memory by dynamic routing algorithm.
- QIM improves the problem of instance-level diversity

Thanks.
And Your Slogan Here.

Speaker name and title

Date