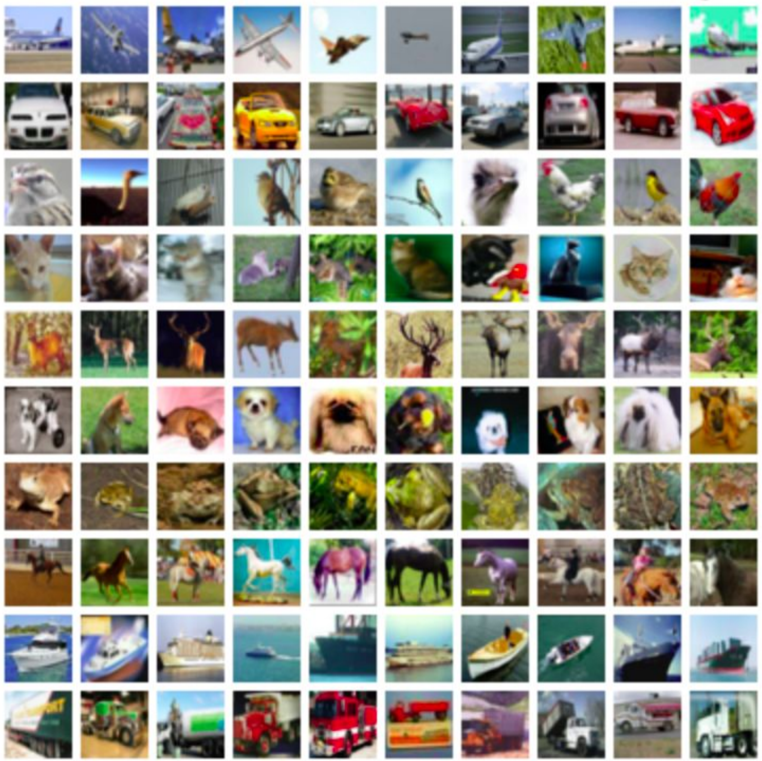
Image Deformation Meta-Networks for One-Shot Learning

- *Zitian Chen et.al*

Presented By
Shivani T, Sarmishta B, Sahana M

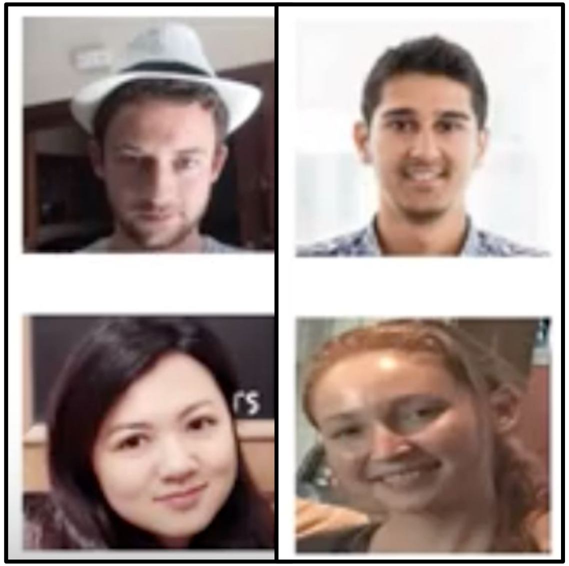
Problem Statement

Traditional Object recognition models use large sets of labelled data to accurately recognize the objects.



One Shot Learning

Learn the object information using **one or few** training image samples



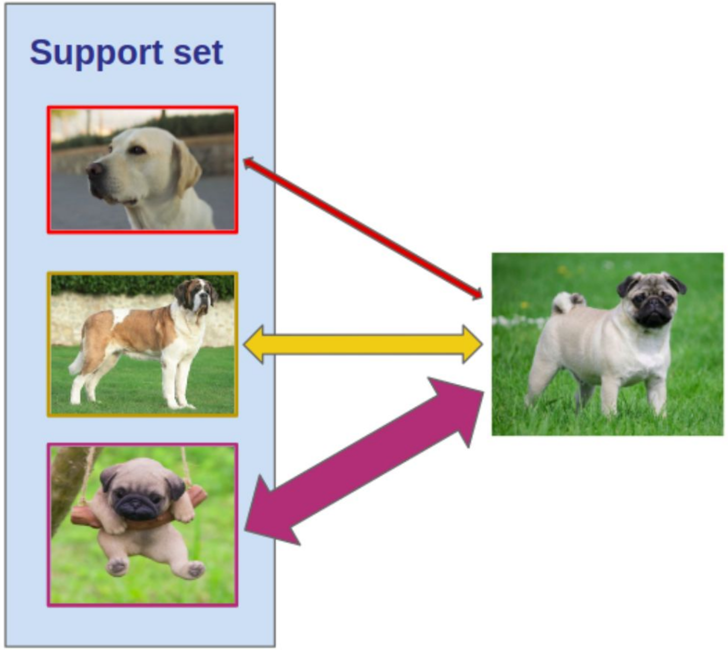
Labelled Data

Novel Data

Related Work

Metric Learning

- Siamese Networks
- Deals with optimizing the metrics.
- Metric: Similarity



Data Augmentation

- Techniques like rotating, cropping, flipping, rescaling, adding noise.
- Synthesising new Images - GANs

Cons:

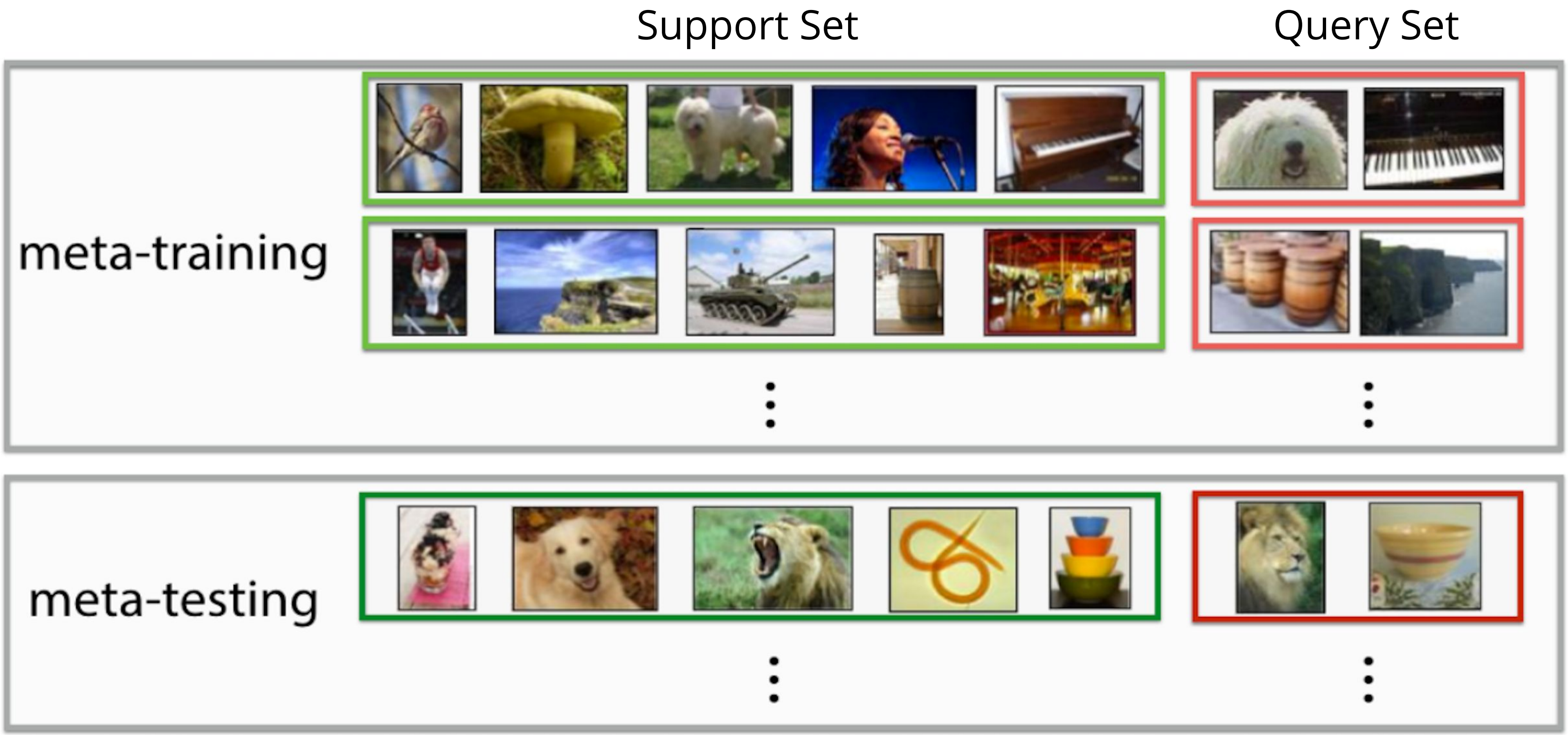
- The visual similarity of the new images is the same
- Doesn’t directly work on increasing one-shot accuracy.

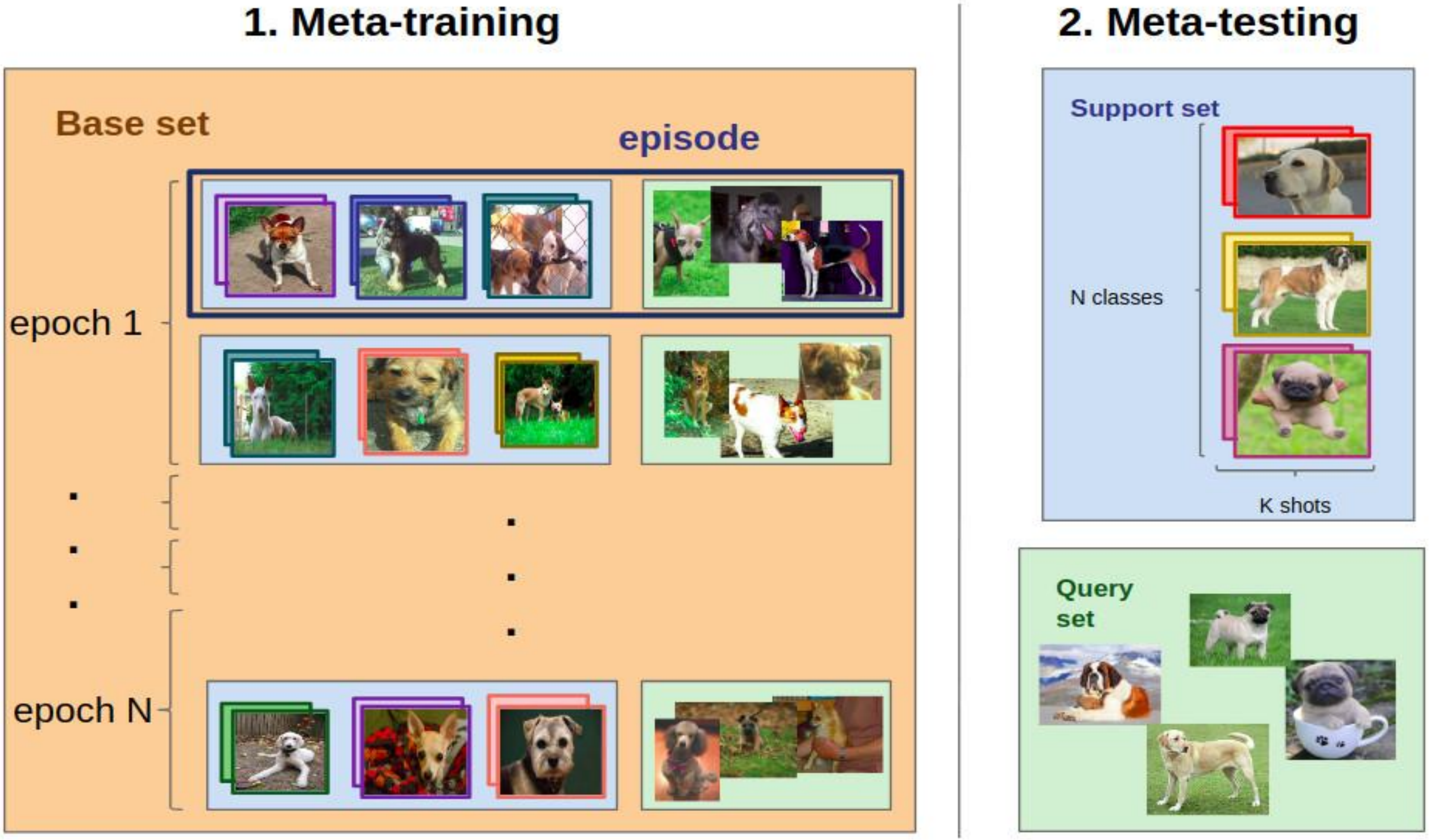
Mixup augmentation

Cons:

- Linear combination, without taking into account structural relation between patches.
- Label of generated image is just linear combination of label vectors.

Meta Learning





1. Meta-training
2. Meta-testing
Base set
episode
epoch 1
epoch N
Support set
N classes
K shots
Query
set

IDeMe-Net

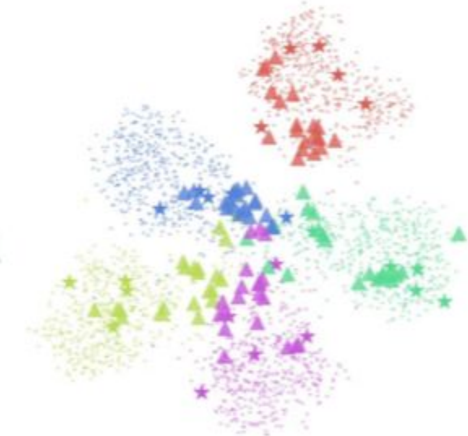

(a)


(b)


(c)


(d)

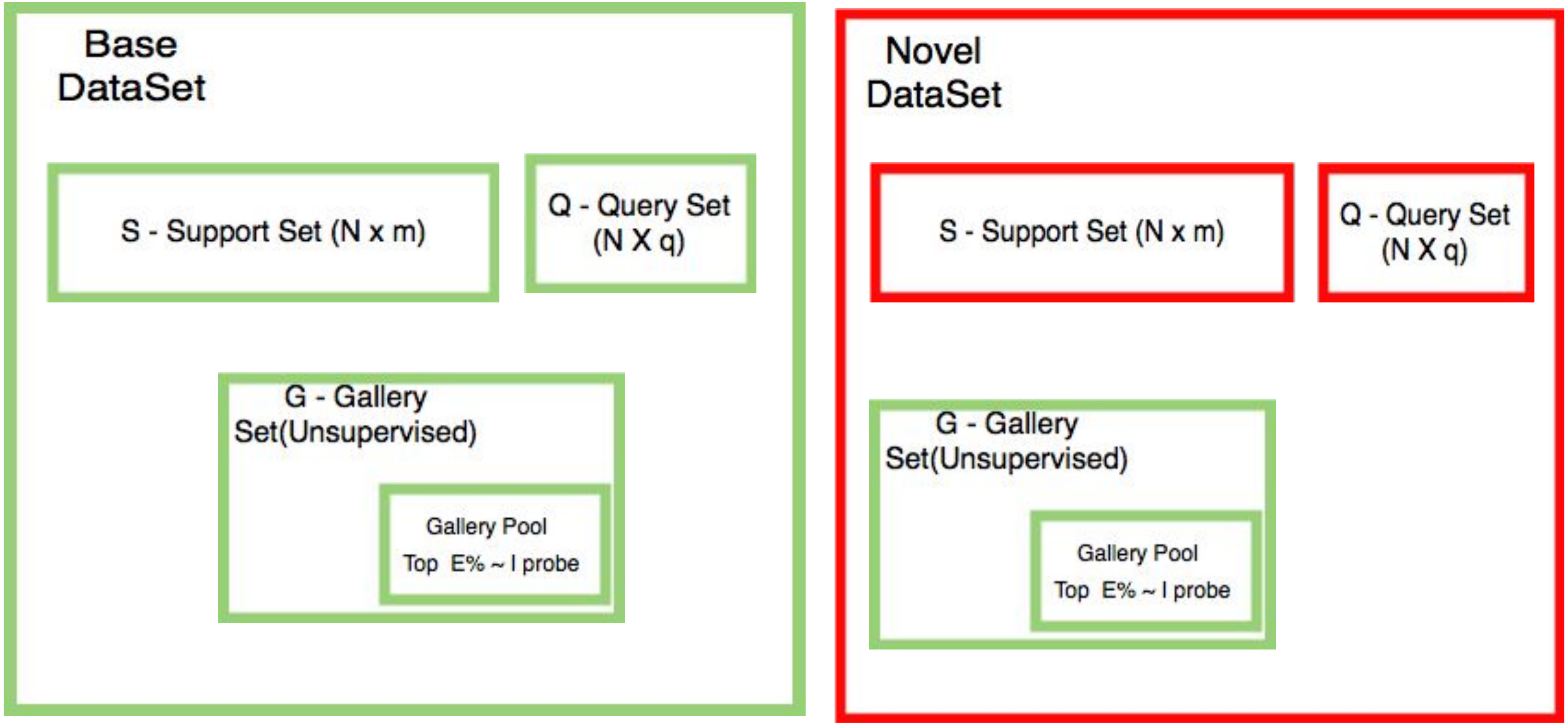

(e)



Contributions - IDeMe-Net

1. Deformation network using meta learning
2. Exploiting the complementarity and interaction between the probe and gallery image patches.
3. Increase the state-of art performance of ImageNet and *mini*Imagenet datasets.

Data Sampling



Model Architecture-Overview

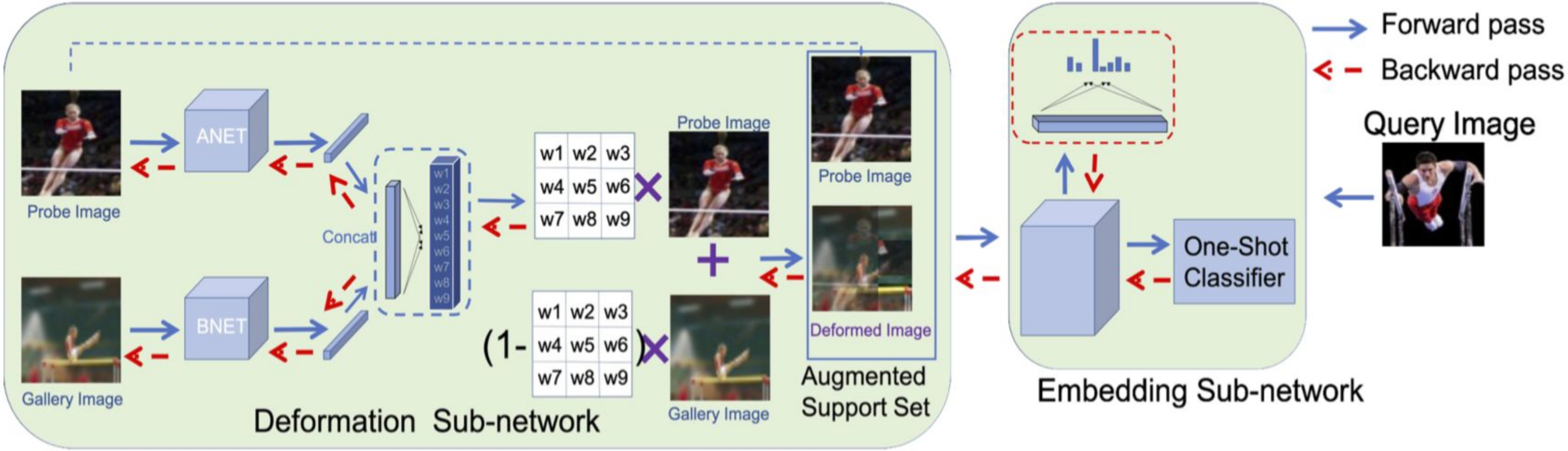
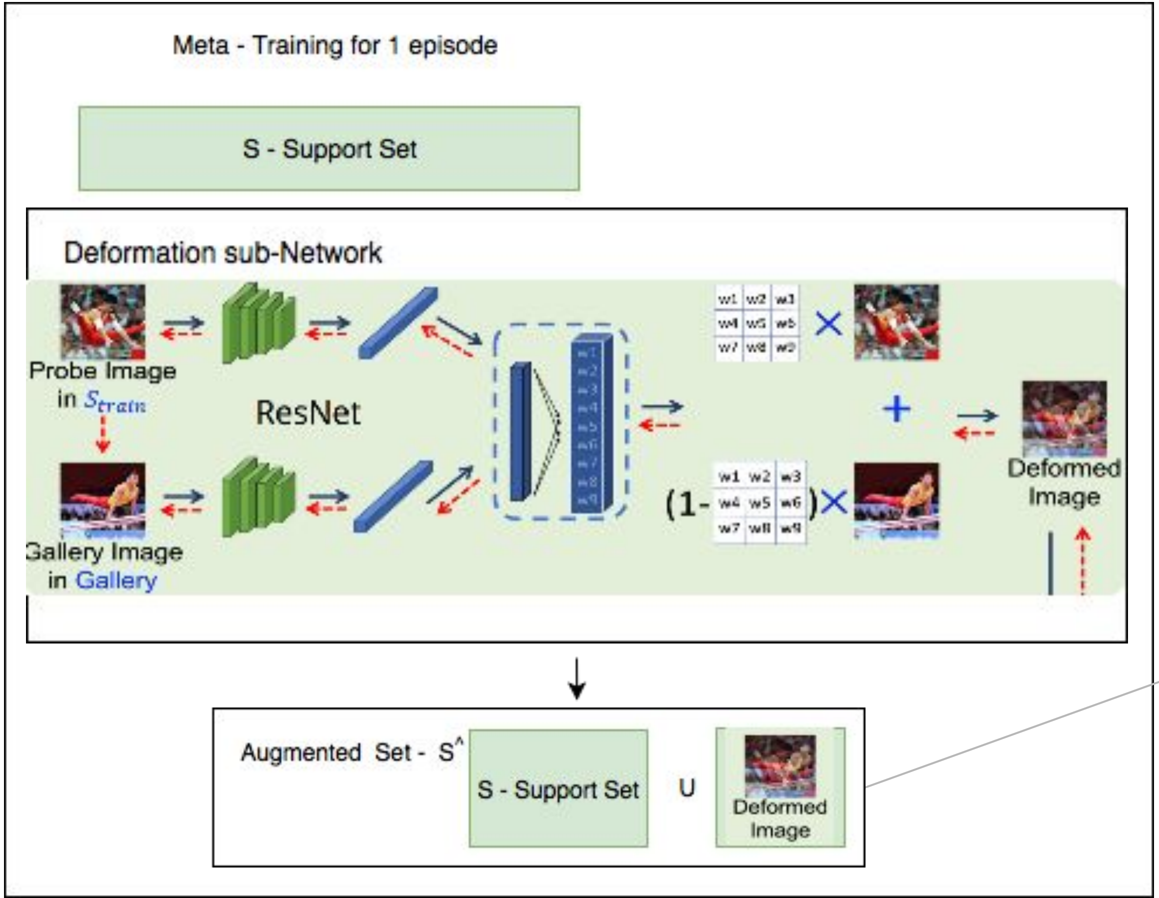


Figure 2. The overall architecture of our image deformation meta-network (IDeMe-Net).

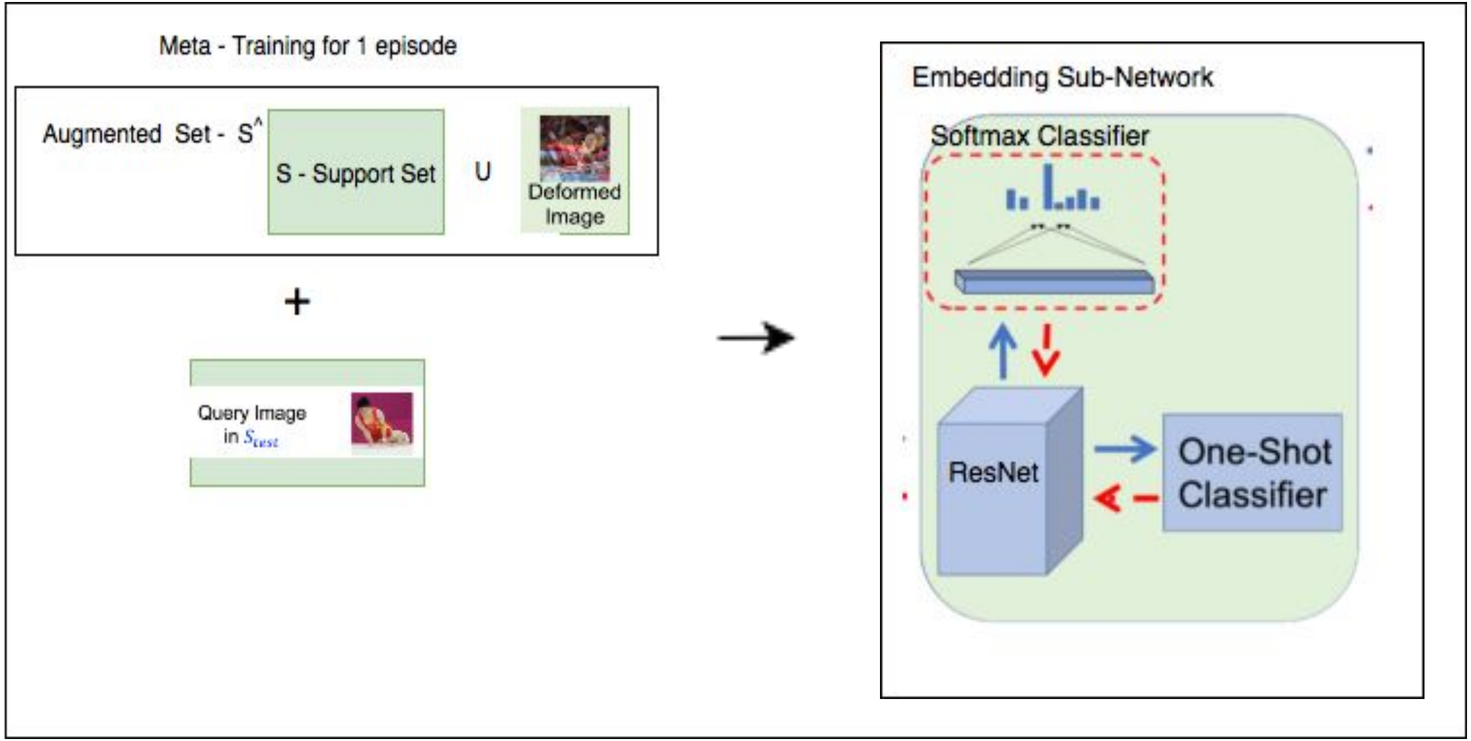
Deformation Sub-Network



Generalization- For any I_{probe} , we sample n_{aug} gallery images and produce n_{aug} synthesized deformed images

$$\left\{\left(\mathbf{I}_{probe}^{i}, y_{probe}^{i}\right), \left\{\left(\mathbf{I}_{syn}^{i,j}, y_{probe}^{i,j}\right)\right\}_{j=1}^{n_{aug}}\right\}_{i=1}^{N\times n}$$

Embedded Sub-Network



Training Losses

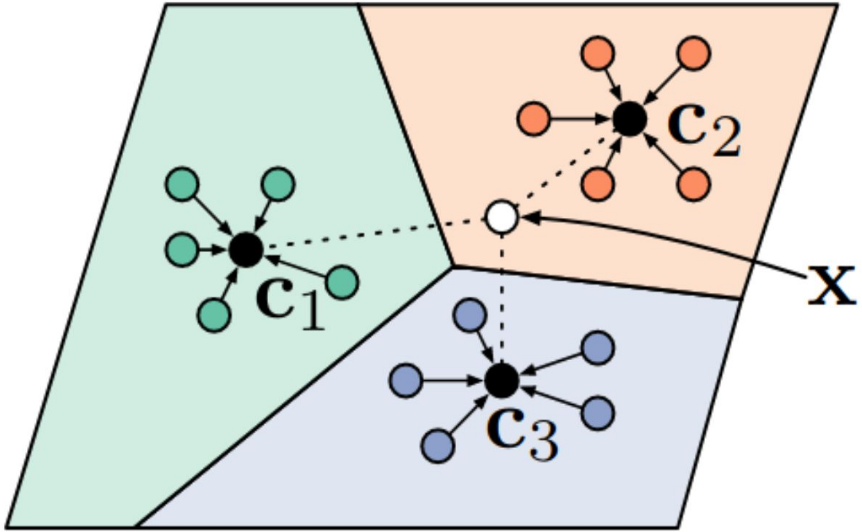
Prototype Loss: $$\min_{\theta}\mathbb{E}_{L\sim D_{base}}\mathbb{E}_{S,G,Q\sim L}\left[\sum_{(\mathbf{I}_i,y_i)\in Q} -\log P_{\theta}\left(y_i \mid \mathbf{I}_i\right)\right],$$

$$P_{\theta}\left(y_i = c|\mathbf{I}_i\right) = \frac{\exp\left(-\left\|f_{\theta_{emb}}\left(\mathbf{I}_i\right) - p_{\theta}^{c}\right\|\right)}{\sum_{j=1}^{N}\exp\left(-\left\|f_{\theta_{emb}}\left(\mathbf{I}_i\right) - p_{\theta}^{j}\right\|\right)},$$

$\|\cdot\|$ *denotes the euclidean distance.*

$$p_{\theta}^{c} = \frac{1}{Z}\sum_{(\mathbf{I}_i,y_i)\in\tilde{S}} f_{\theta_{emb}}\left(\mathbf{I}_i\right)\cdot\left[\!\left[y_i = c\right]\!\right],$$

Prototype vector p_c for each class in S:



Putting it all together - Training end-to-end

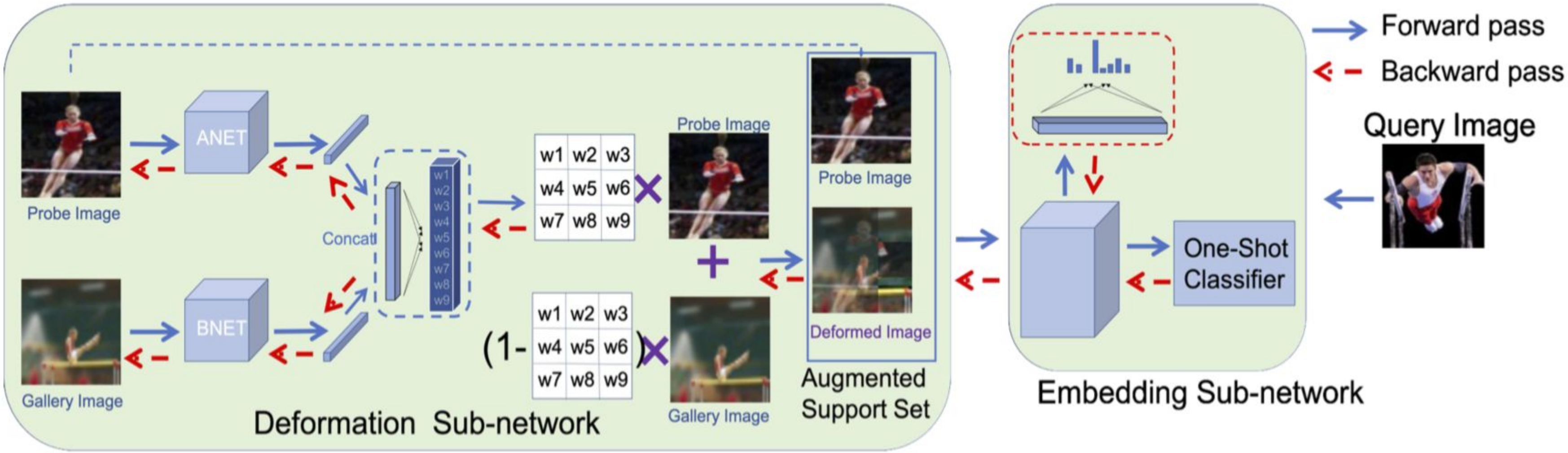


Figure 2. The overall architecture of our image deformation meta-network (IDeMe-Net).

Algorithm

Algorithm 1 Meta-training procedure of our IDeMe-Net f_θ. G is the fixed gallery constructed from $\mathcal{C}_{base}$.

procedure META-TRAIN_EPISODE
The procedure of one meta-training episode

Random sampling for episodes

  $L \leftarrow$ randomly sample N classes from $\mathcal{C}_{base}$
  $S \leftarrow$ randomly sample instances belonging to L
  //sample the support set
  $Q \leftarrow$ randomly sample instances belonging to L
  //sample the query set

Learn the feature maps for S

  train the prototype classifier P from $f_{\theta_{emb}}(S)$

Generate the Pool from the gallery.

  $\tilde{S} \leftarrow S$ ▷ initialize the augment support set
  for c in L **do** ▷ enumerate the chosen classes
    pool $\leftarrow$use P to select $\epsilon\%$ images in G that have the highest class probability of c
    for $(\mathbf{I}_{prob}, c)$ in S_c **do**
      for $j = 1$ to n_{aug} **do**
        $\mathbf{I}_{gallery} \leftarrow$ randomly sample instances from *pool*

Synthesise n_{aug} Images per prob

        $\mathbf{I}_{syn} \leftarrow f_{\theta_{def}}(\mathbf{I}_{prob}, \mathbf{I}_{gallery})$
        $\tilde{S} \leftarrow \tilde{S} \cup (\mathbf{I}_{syn}, c)$
      end for
    end for
  end for

Train and Propagate Loss

  train the prototype classifier $\tilde{P}$ from $f_{\theta_{emb}}(\tilde{S})$
  use $\tilde{P}$ to classify $f_{\theta_{emb}}(Q)$ and obtain the prototype loss
  use the softmax classifier to classify $f_{\theta_{emb}}(\tilde{S})$ and obtain the CELoss
  update θ_{emb} with the CELoss
  update θ_{def} with the prototype loss
end procedure

Experimental Setup

2 Benchmarks:

- imageNet 1K Challenge
- miniImageNet

- First 20 epochs, train one sub-network at a time
- Every 30 epochs, reduce learning rate
- End to end convergence after 100 epochs

Image 1K Challenge Dataset

	Method	$m = 1$	2	5	10	20
Baselines	Softmax	– /16.3	– /35.9	– /57.4	- / 67.3	– /72.1
	LR	18.3/42.8	26.0/54.7	35.8/66.1	41.1/71.3	44.9/74.8
	SVM	15.9/36.6	22.7/48.4	31.5/61.2	37.9/69.2	43.9/74.6
	Prototype Classifier	17.1/39.2	24.3/51.1	33.8/63.9	38.4/69.9	44.1/74.7
Competitors	Matching Network [28]	– / 43.0	– / 54.1	– / 64.4	– / 68.5	– / 72.8
	Prototypical Network [22]	16.9/41.7	24.0/53.6	33.5/63.7	37.7/68.2	42.7/72.3
	Generation-SGM [9]	– / 34.3	– / 48.9	– / 64.1	– / 70.5	– /74.6
	PMN [30]	– / 43.3	– / 55.7	– / 68.4	– / 74.0	– / 77.0
	PMN w/ H [30]	– / 45.8	– / 57.8	– / 69.0	– / 74.3	– / 77.4
	Cos & Att. [8]	– / 46.0	– / 57.5	– / 69.1	– / **74.8**	– / **78.1**
	CP-AAN [6]	– / 48.4	– / 59.3	– / 70.2	– / 76.5	– / 79.3
Augmentation	Flipping	17.4/39.6	24.7/51.2	33.7/64.1	38.7/70.2	44.2/74.5
	Gaussian Noise	16.8/39.0	24.0/51.2	33.9/63.7	38.0/69.7	43.8/74.5
	Gaussian Noise (feature level)	16.7/39.1	24.2/51.4	33.4/63.3	38.2/69.5	44.0/74.2
	Mixup [36]	15.8/38.7	24.6/51.4	32.0/61.1	38.5/69.2	42.1/72.9
Ours	IDeMe-Net	**23.1/51.0**	**30.1/60.9**	**39.3/70.4**	**42.7**/73.4	**45.0**/75.1

Table 1. **Top-1 / Top-5 accuracy (%) on novel classes of the ImageNet 1K Challenge dataset.** We use **ResNet-10** as the embedding sub-network. m indicates the number of training examples per class. Our IDeMe-Net consistently achieves the best performance.

Image 1K Challenge Dataset

Method	$m = 1$	2	5	10
Softmax	– /28.2	– /51.0	– / 71.0	– /78.4
SVM	20.1/41.6	29.4/57.7	42.6/72.8	49.9/79.1
LR	22.9/47.9	32.3/61.3	44.3/73.6	50.9/78.8
Proto-Clsf	20.8/43.1	29.9/58.1	42.4/72.3	49.5/79.0
G-SGM [9]	– /47.3	– /60.9	– /73.7	– /79.5
PMN [30]	– / 53.3	– / 65.2	– / 75.9	– / 80.1
PMN w/ H [30]	– / 54.7	– / 66.8	– / **77.4**	– / **81.4**
IDeMe-Net (Ours)	**30.3/60.1**	**39.7/69.6**	**47.5/77.4**	**51.3**/80.2

Table 2. **Top-1 / Top-5 accuracy (%) on novel classes of the Imagenet 1K Challenge dataset.** We use **ResNet-50** as the embedding sub-network. m indicates the number of training examples per class. Proto-Clsf and G-SGM denote the prototype classifier and generation SGM [9], respectively.

Image 1K Challenge Dataset - Ablation Study

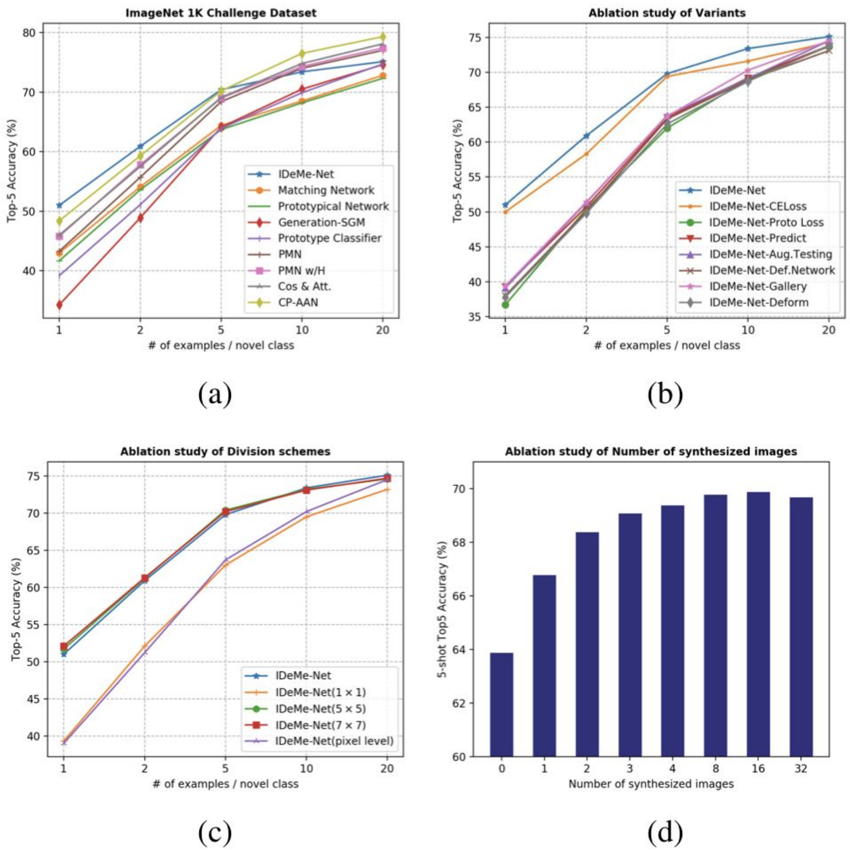


Figure 3. **Ablation study on ImageNet 1K Challenge dataset:**

Ablation Study Conclusions

1. Using cross-entropy loss to update the embedding sub network & prototype loss to update the deformation network achieves the best performance

2. Selecting the gallery images is the key to diversifying the deformed images

3. Performance improvements can mainly be attributed to the diversified deformed images not the embedding sub-network

4. Deformation sub-network effectively exploits the relationship between the probe and gallery image patches

	Method	$m = 1$	2	5	10	20
Baselines	LR	18.3/42.8	26.0/54.7	35.8/66.1	41.1/71.3	44.9/74.8
	Prototype Classifier	17.1/39.2	24.3/51.1	33.8/63.9	38.4/69.9	44.1/74.7
Variants	IDeMe-Net - CELoss	21.3/50.0	28.0/58.3	37.7/69.4	41.3/71.6	44.3/74.3
	IDeMe-Net - Proto Loss	15.3/36.7	21.4/50.4	31.7/62.0	37.9/69.0	43.7/73.7
	IDeMe-Net - Predict	17.0/39.3	24.0/50.7	33.6/63.5	38.0/69.2	43.7/73.8
	IDeMe-Net - Aug. Testing	17.0/39.1	24.30/51.3	33.5/63.8	38.0/69.1	43.8/74.5
	IDeMe-Net - Def. Network	15.9/38.0	24.1/50.1	32.6/63.3	38.2/68.9	42.4/73.1
	IDeMe-Net - Gallery	17.5/39.4	24.2/51.4	33.5/63.7	38.7/70.3	44.4/74.5
	IDeMe-Net - Deform	15.7/37.8	22.7/49.8	31.9/62.6	38.0/68.7	43.5/73.8
Patch Size	IDeMe-Net (1×1)	16.2/39.3	24.4/52.1	32.9/63.0	38.8/69.5	42.7/73.2
	IDeMe-Net (5×5)	**24.1**/51.7	30.3/61.2	**39.6/70.4**	42.4/73.2	44.3/74.6
	IDeMe-Net (7×7)	23.8/**52.1**	30.2/**61.3**	39.1/70.2	**42.7**/73.1	44.5/74.7
	IDeMe-Net (pixel level)	17.3/39.0	23.8/51.2	34.1/63.7	38.5/70.2	43.9/74.5
Ours	IDeMe-Net	23.1/51.0	**30.4**/60.9	39.3/70.4	**42.7/73.4**	**45.0/75.1**

Table 3. **Top-1 / Top-5 accuracy (%) of the ablation study on novel classes of the ImageNet 1K Challenge dataset.** We use ResNet-10 as the embedding sub-network. m indicates the number of training examples per class. Our full model achieves the best performance.

miniImageNet Dataset

Method	*mini*ImageNet (%)	
	1-shot	5-shot
MAML [5]	48.70±1.84	63.11±0.92
Meta-SGD [13]	50.47±1.87	64.03±0.94
Matching Network [28]	43.56±0.84	55.31±0.73
Prototypical Network [22]	49.42±0.78	68.20±0.66
Relation Network [23]	57.02±0.92	71.07±0.69
SNAIL [14]	55.71±0.99	68.88±0.92
Delta-Encoder [21]	58.7	73.6
Cos & Att. [8]	55.45±0.89	70.13 ±0.68
Prototype Classifier	52.54±0.81	72.71±0.73
IDeMe-Net (Ours)	**59.14**±0.86	**74.63**±0.74

Table 4. **Top-1 accuracy (%) on novel classes of the *mini*ImageNet dataset.** "±" indicates 95% confidence intervals over tasks.

Conclusion

- Simple Approach
- Deformation Images being the major contribution.
- Extensive ablation study

- Extend this augmentation technique to more complex networks.
- Extend to KNN classifier

Questions?