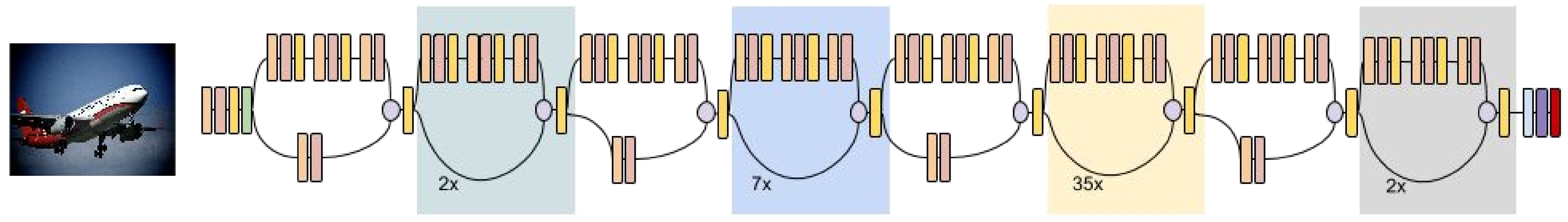
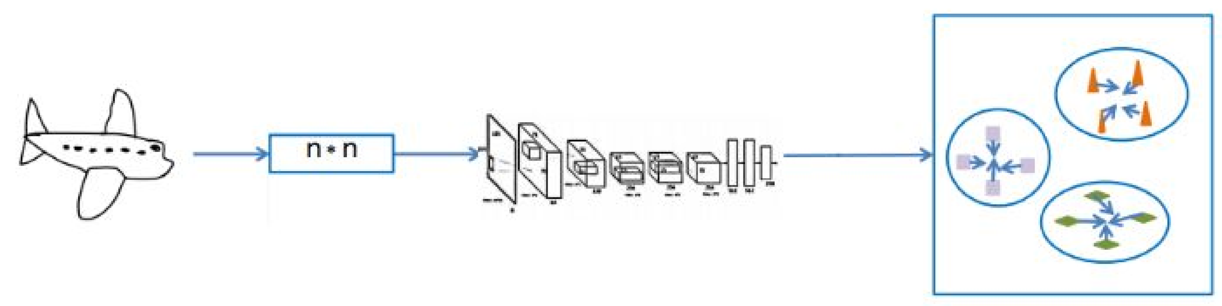
SwiDeN : Convolutional Neural Networks For Depiction Invariant Object Recognition

[http://val.serc.iisc.ernet.in/]
Ravi Kiran Sarvadevabhatla*, Shiv Surya*, Srinivas S S Kruthiventi, R. Venkatesh Babu (*equal contribution)
Video Analytics Lab, Indian Institute of Science, Bangalore, INDIA

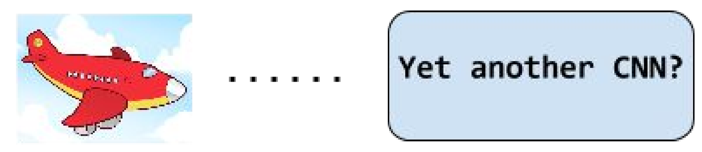
CURRENT CLASSIFICATION PARADIGM



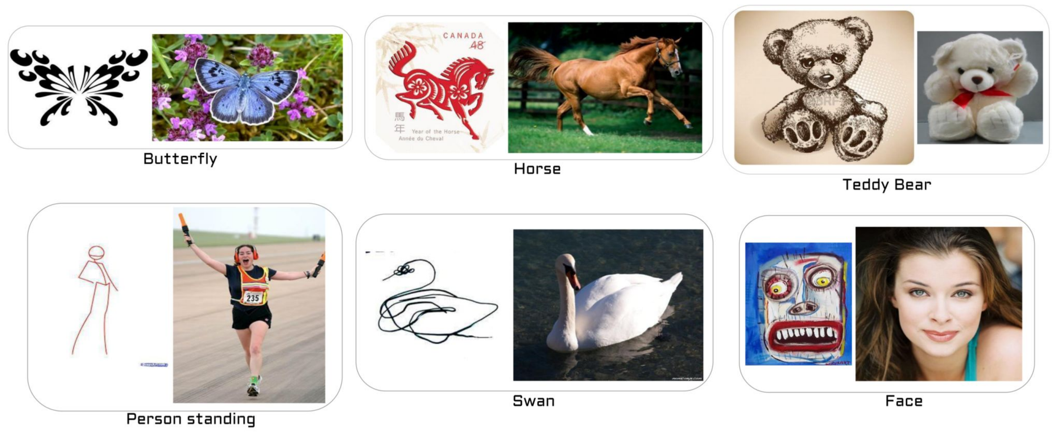
ResNet-152[1]



Sketch-a-Net[2]



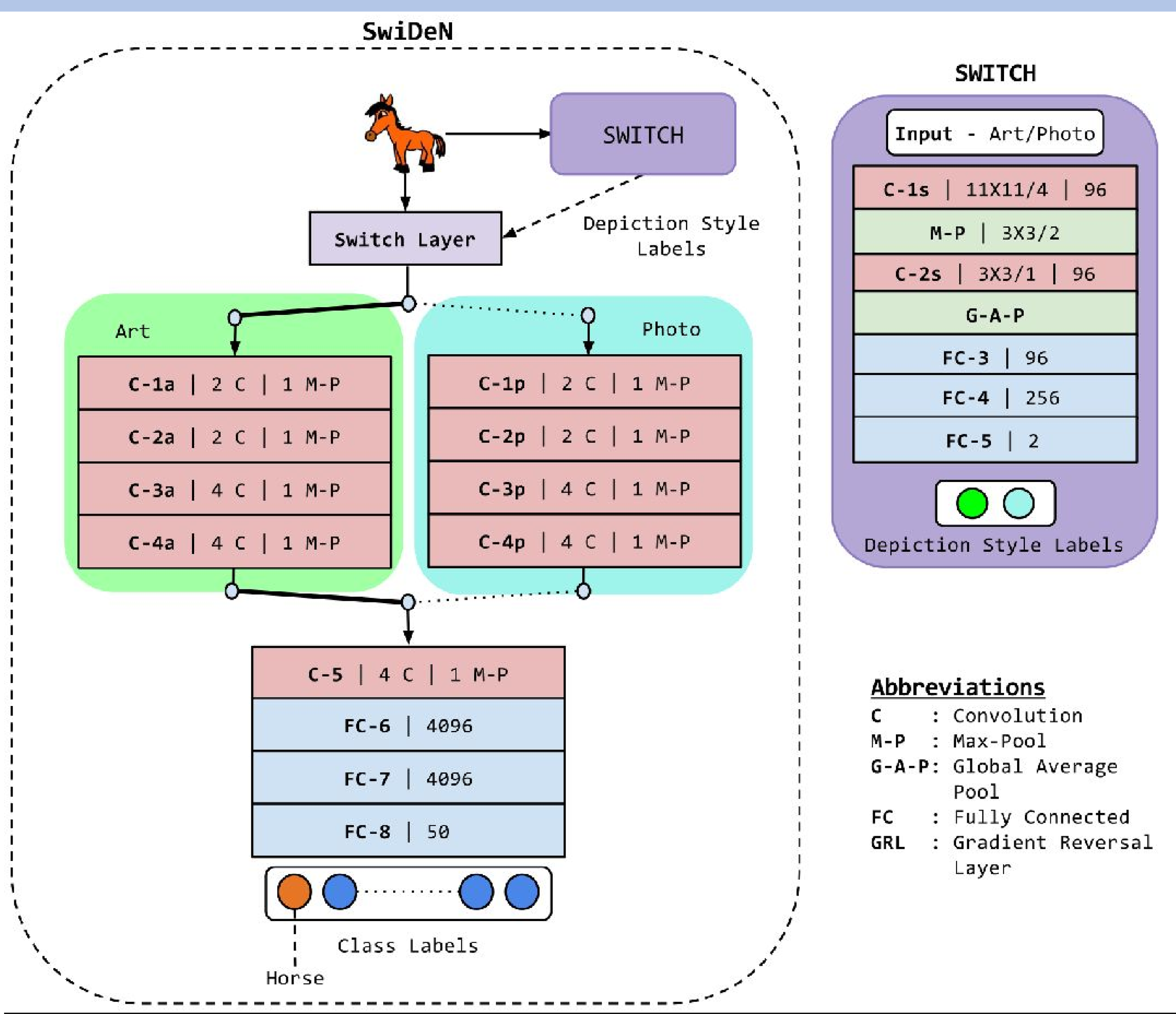
DEPICTION INVARIANT OBJECT RECOGNITION : CHALLENGES…



Dataset: Photo-Art-50

- 90-138 images per class.
- Approximately half photo and half art images.
- Contains objects depicted in multiple styles.

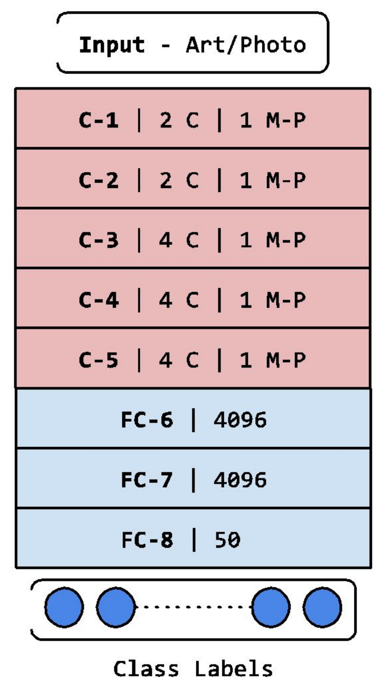
OUR APPROACH : SWITCHING DEEP NETWORK



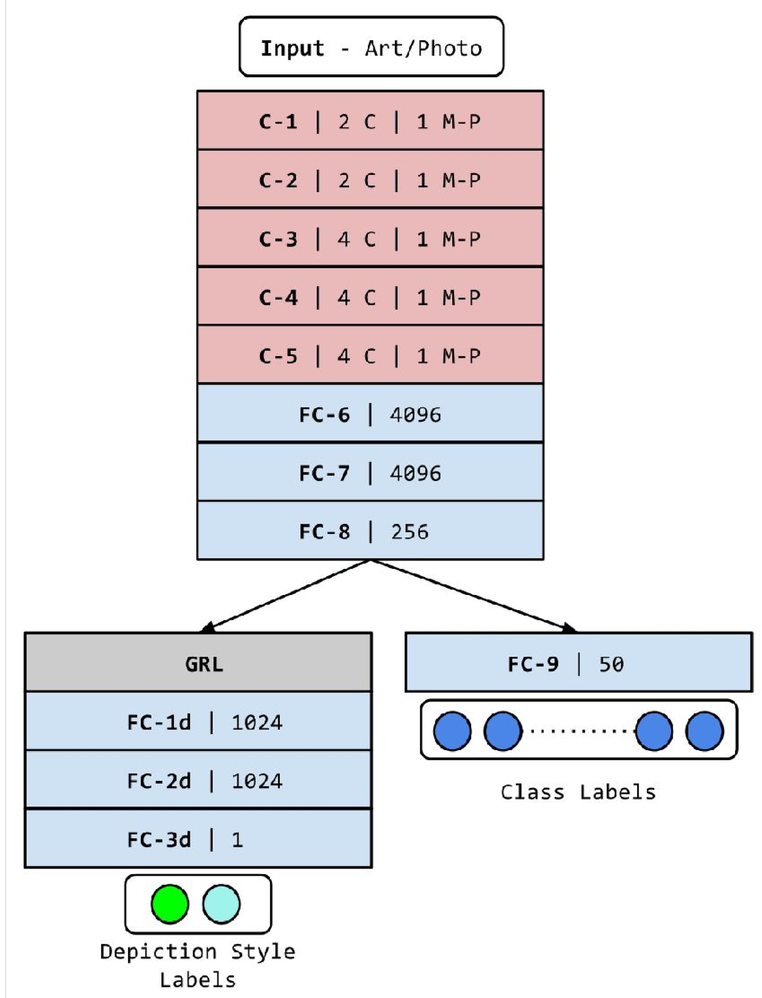
SwiDeN: Switching Deep Network

- Novel `deep' depictive style-based switching mechanism.
- Set of deep sub-networks exist for each depictive style.
- Final set of common layers learn depiction-invariant representation.
- Custom **SWITCH** network serve as a relay mechanism between the initial depiction-specific sub-networks and the shared, deeper depiction-invariant layers.
- **SwiDeN** and **SWITCH** are trained using transfer learning with batch stochastic gradient descent.
- Code available at https://github.com/val-iisc/swiden

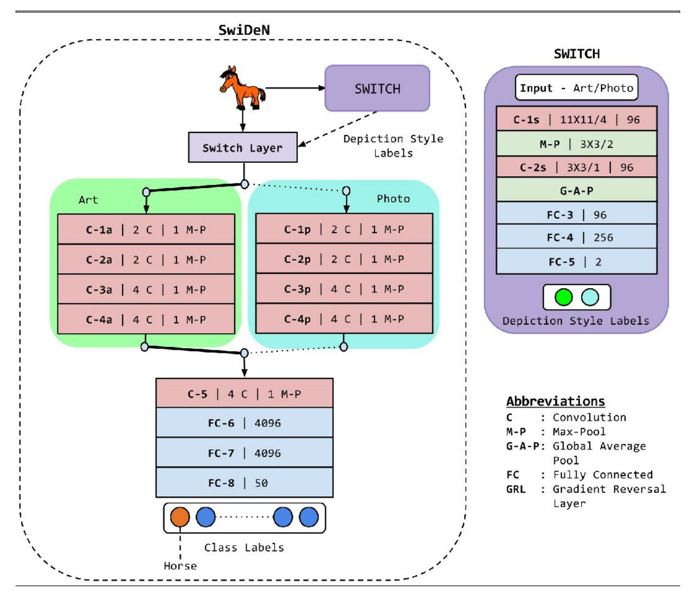
OTHER APPROACHES



Baseline



Gradient Reversal Network (GRN)



SwiDeN

RESULTS

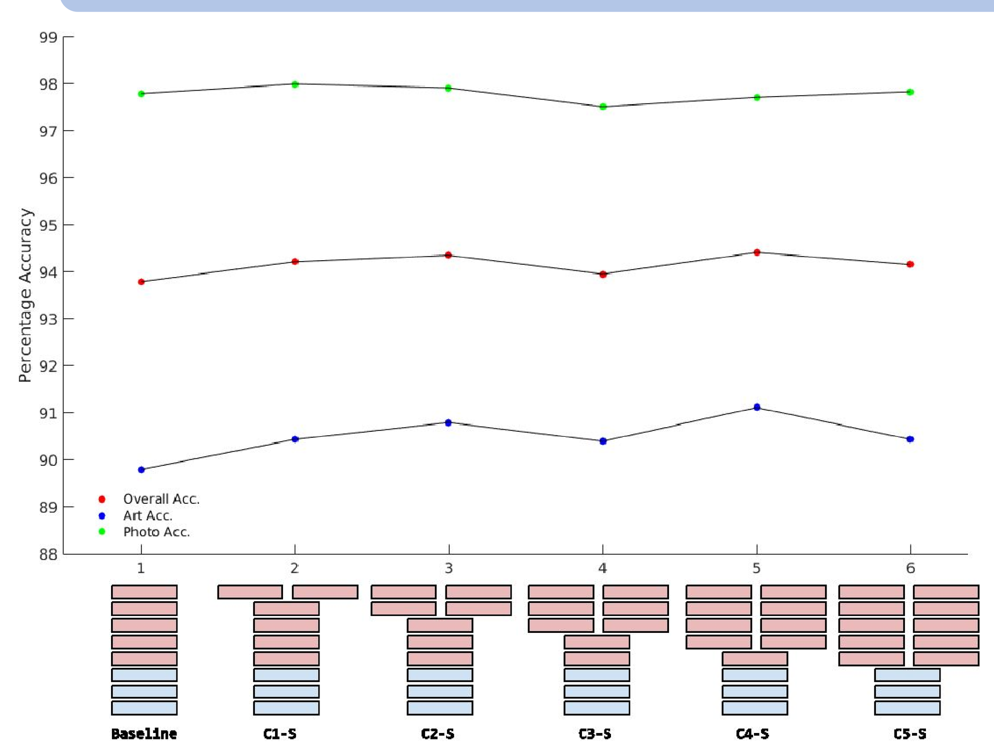
Arch.	Overall Acc.	Art Acc.	Photo Acc.
Baseline	93.80%	89.80%	**97.80%**
GRN	92.64%	88.52%	96.76%
SwiDeN(Ours)	**94.42%**	**91.12%**	97.72%

Table 1: Classification accuracy for different architectures.

Arch.	Overall Acc.	Art Acc.	Photo Acc.
Wu et al.[7]	89.67%	**89.06%**	90.29%
SwiDeN (Ours)	**93.02%**	88.47%	**97.56%**

Table 2: Classification accuracy on train-test splits by Cai et al. [3].

ANALYSIS



Average classification accuracy for different **SwiDeN** architectures

- As depth of depiction sub-networks increase, we observe that the general classification accuracy and art classification accuracy shows an upward trend.

Some example images that are misclassified by **SWITCH**.

- **SWITCH** achieves an average accuracy of **83.7%** (80.6% for 'Art' and 86.8% for 'Photo') .
- **SWITCH's** inability to achieve 100% accuracy can be attributed to the fact that some photo images have a predominantly artistic quality and vice-versa.

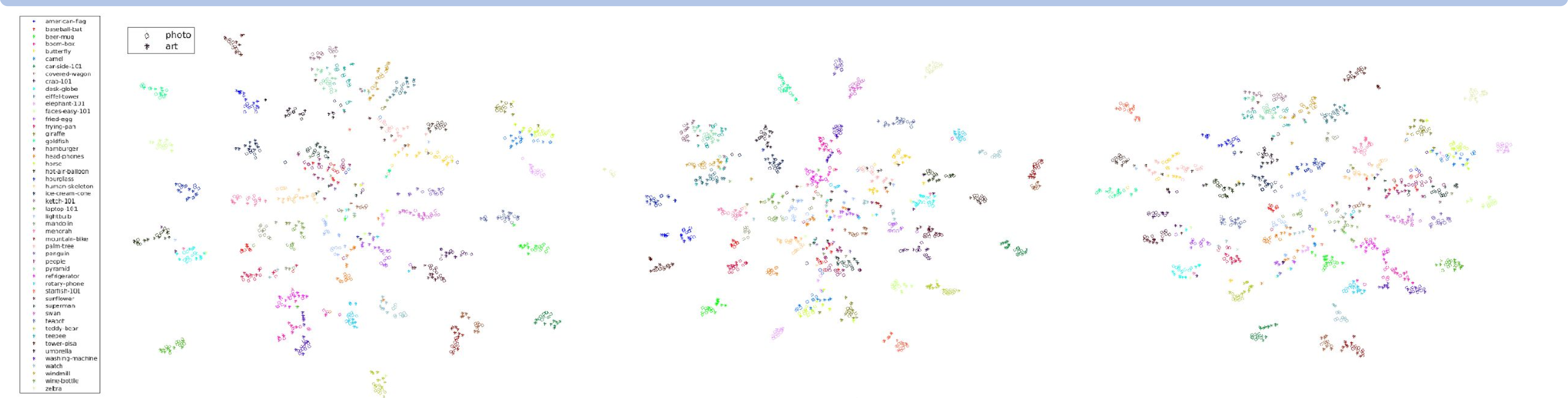
REFERENCES

[1] He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).

[2] Yu, Qian, et al. "Sketch-a-net that beats humans." *arXiv preprint arXiv:1501.07873* (2015).

[3] Wu et al. Learning graphs to model visual objects across different depictive styles, ECCV 2014.

T-SNE VISUALIZATIONS



SwiDeN

Baseline

Gradient Reversal Network (GRN)