Siamese Networks

Alan Yuille

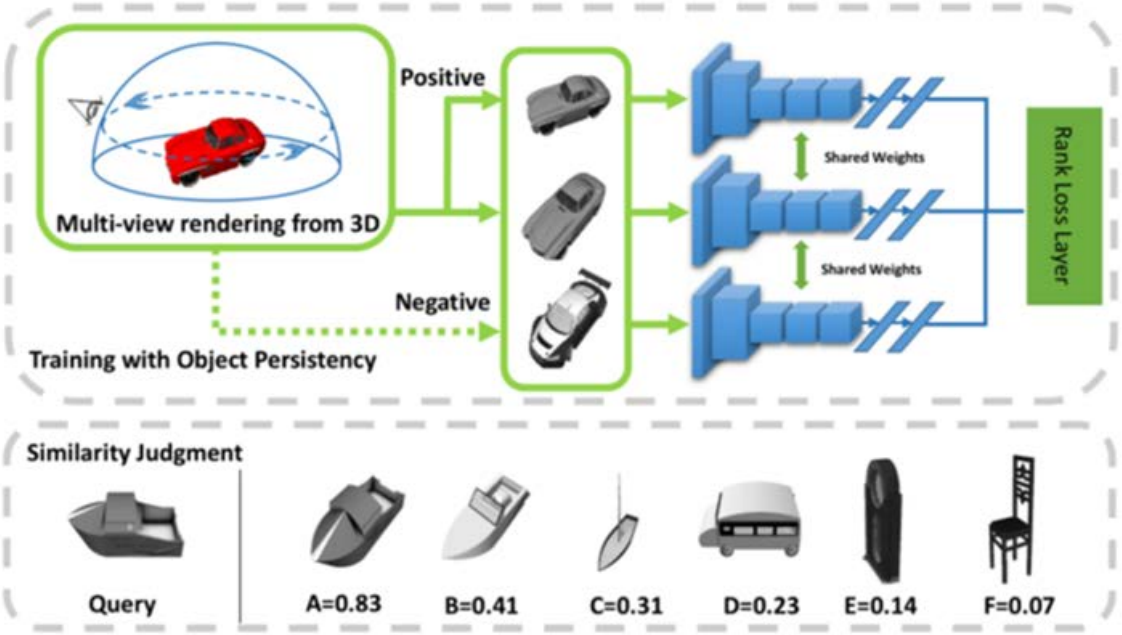
Similarity Perception of Novel Objects.

- Humans can perform similarity judgments on novel objects. E.g, Tufa's.
- Humans learn from image sequences -- i.e., observe an object from several viewpoints and know that it is the same, even if we do not know what its name is. Object Persistence (OP).
- Our goal: use a Deep Net – specifically a Siamese-Triplet Net – to learn similarity judgments. Train on different views of the same object, test on novel objects.
- Note – it is very hard to find objects that adults, or even young children, have never seen before. Presumably a lot of learning can bootstrap from known objects. Why researchers in the 90's had to test on paperclips.
- X. Lin et al. ICLR 2017.

Siamese-Triplet Nets

- These were developed to perform similarity judgments – originally for signature verification (Bromley et al. 1993).
- Researchers have used them, on image sequences, to learn Deep Net features without class label supervision (e.g., Wang & Gupta 2015).
- Siamese-Triplet Nets consist of three Deep Nets which combine to give a binary result – similar or non-similar.
- We train ours using object persistence – an image sequence gives us different views of the same object – so we call it OPnet.
- We train on known objects and test on unknown objects.

Siamese-Triplet Network



- Training (upper panel) and testing (lower panel).
- The lower panel shows similarity scores given by our OPnet.
- Different views of the same object are the most similar, followed by different objects in the same category, and finally objects belonging to different categories.

Training Data

- Train on a subset of ShapeNet (Chang et al. 2015). These are 3D object models, e.g., cars and chairs, which are rendered from different viewpoints.
- Select 7,000 3D object models belonging to 55 categories. For each model, render 12 different views by rotating the cameras along the equator from a 30◦ elevation angle and taking photos of the object at 12 equally separated azimuthal angles.
- For training, sample 200 object models from 29 categories of ShapeNet.

Testing Data

- We test on novel objects from ShapeNet, Pokemon, Synthetic, & Tufa's. This tests transfer to objects which have not been seen before.
- Novel instance: Created by rendering additional 20 novel objects from each of the 29 categories used in training the OPnet.
- Novel category: Created by rendering objects from 26 untrained categories. This is a more challenging test of the transfer of view-manifold learning to novel categories.
- Pokemon. 438 CAD models of Pokemon from an online database.
- Synthetic. Textureless objects with completely novel shapes. The dataset consists of 5 categories, with 10 instances for each category.
- Tufa's. Objects from Tenenbaum et al. (2011), where ground truth is based on human similarity judgments.

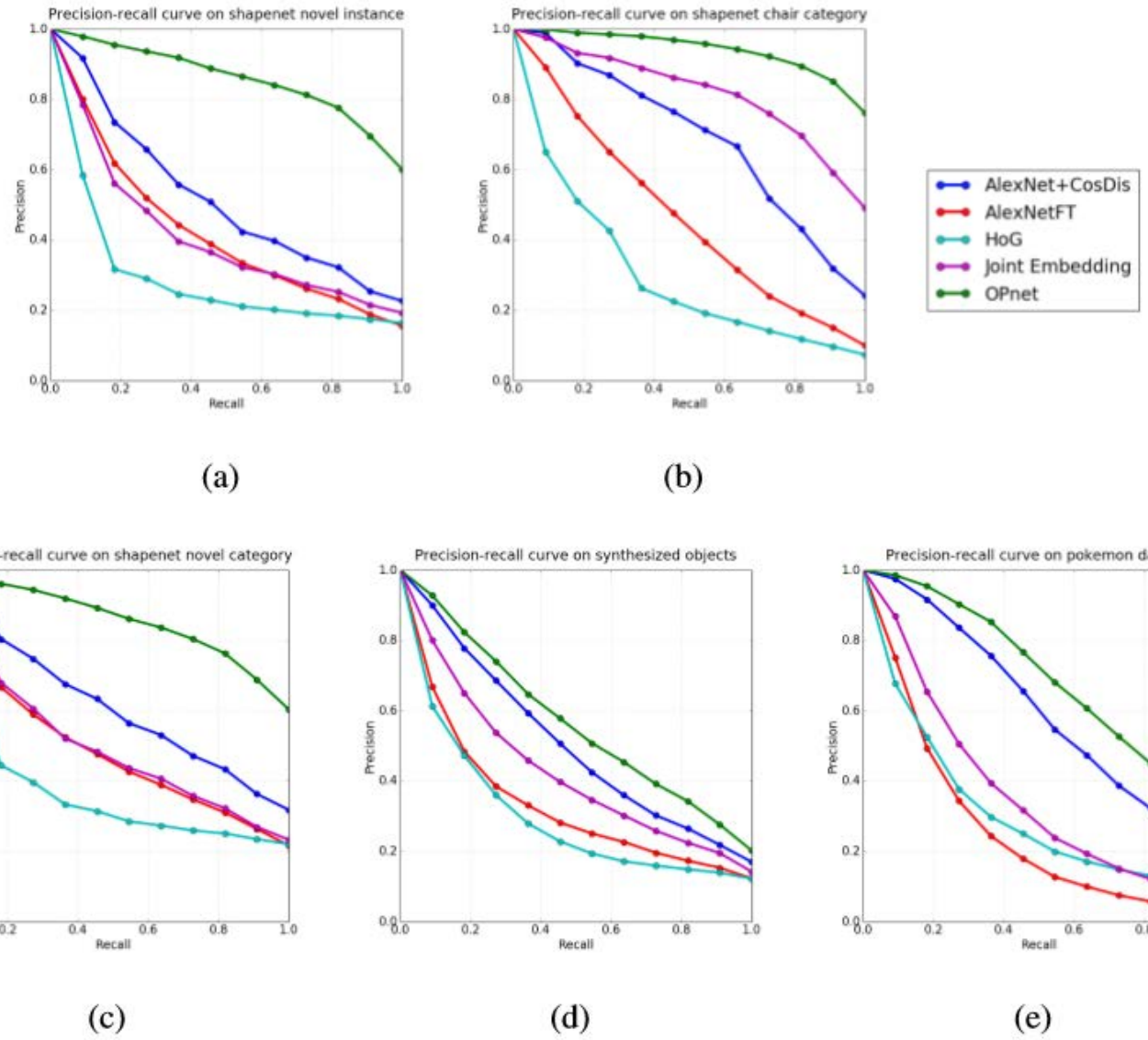
Findings: Object Retrieval

- Similarity Learning transfers across datasets.
- In the object instance retrieval task, for each image P containing object O of category C in the test set, the network is asked to rank all other images in C, such that images for O should have higher similarity score than images for other objects in C.

	Novel instance	Novel category	Synthesized objects	Pokemon	Chair
HoG	0.316	0.391	0.324	0.332	0.322
AlexNetFT	0.437	0.503	0.356	0.287	0.478
AlexNet+CosDis	0.529	0.623	0.517	0.607	0.686
AlexNet+EucDis	0.524	0.617	0.514	0.591	0.677
OPnet	**0.856**	**0.855**	**0.574**	**0.697**	**0.938**
Joint-embedding	0.429	0.513	0.443	0.387	0.814

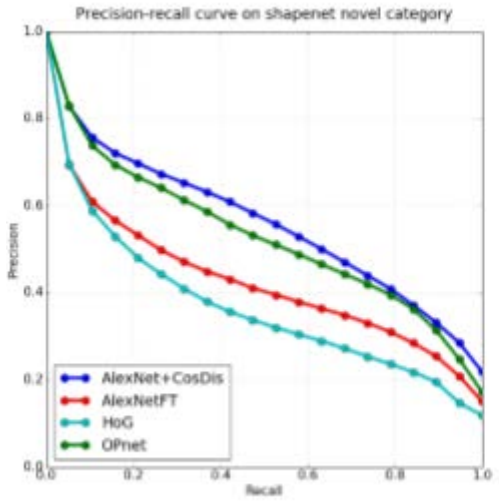
Findings:

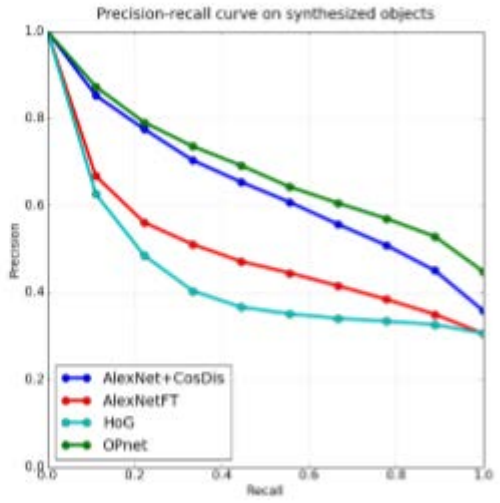
- Comparisons



Findings: Novel Categories.

- Deep Net has the advantage of category knowledge.
- But OPnet does almost as well despite not using category knowledge.

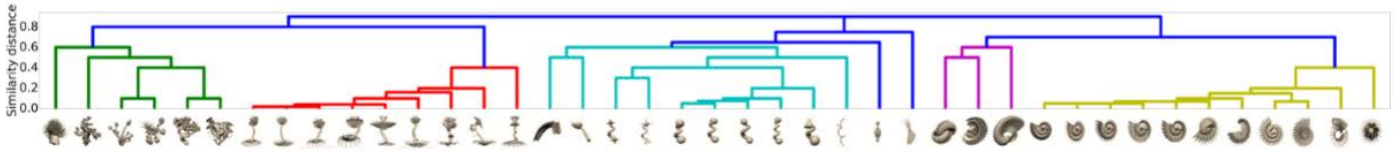




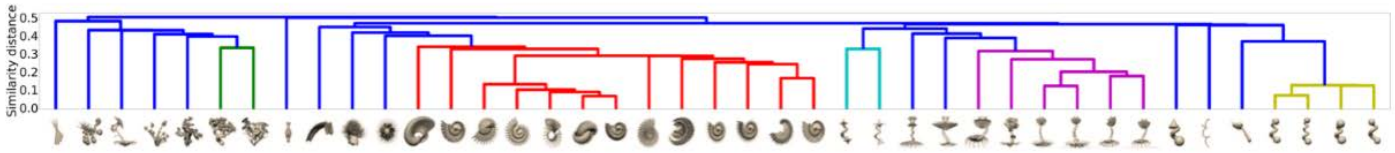
Comparison to Human Similarity Judgments

- Using the novel objects from Tenenbaum et al. (2011), we are able to compare our networks with human similarity perception. We collect 41 images from the paper, one image per object.
- A pairwise similarity matrix is calculated based on the cosine distance of their feature representations. We then perform hierarchical agglomerative clustering to obtain a tree structure, merging the two clusters with shortest distance successively to construct the tree.
- Compare the results with human perception. And with hierarchical clustering using AlexNet features.

Findings: Comparison with Humans.



(a) Grouping by Human Perception



(b) Grouping by AlexNet Features

(c) Grouping by OPnet Features

- Hierarchical clustering of the alien objects, based on (a) human perception, (b) AlexNet features and (c) OPnet features.
- Spearman Correlation: 0.460 with AlexNet, 0.659 with OPnet.

Conclusion

- The Siamese-Triplet network – OPnet –exploits the object persistence constraint and shows transference of similarity judgments to novel objects.
- OPnets performs well, significantly outperforming AlexNet, and also performs well when compared to human perception judgments.

- It seems plausible that object persistence – learning an object using different viewpoints – is an effective way of learning that has some biologically plausibility.