Towards Light-weight Annotations: Fuzzy Interpolative Reasoning for Zero-shot Image Classification

Yang Long[1]
yang.long@ieee.org
Yao Tan[2]
yao.tan@northumbria.ac.uk
Daniel Organisciak[2]
d.organisciak@gmail.com
Longzhi Yang[2]
longzhi.yang@northumbria.ac.uk
Ling Shao[3]
ling.shao@ieee.org

[1] Open Lab, School of Computing Newcastle University, UK
[2] Department of Computer Science and Digital Technologies, Northumbria University, UK
[3] Inception Institute of Artificial Intelligence, UAE

Abstract

Despite the recent popularity of Zero-shot Learning (ZSL) techniques, existing approaches rely on ontological engineering with heavy annotations to supervise the transferable attribute model that can go across seen and unseen classes. Moreover, existing cross-sourcing, expert-based, or data-driven attribute annotations (e.g. Word Embeddings) cannot guarantee sufficient description to the visual features, which leads to significant performance degradation. In order to circumvent the expensive attribute annotations while retaining the reliability, we propose a Fuzzy Interpolative Reasoning (FIR) algorithm that can discover inter-class associations from light-weight *Simile* annotations based on visual similarities between classes. The inferred representation can better bridge the visual-semantic gap and manifest state-of-the-art experimental results.

1 Introduction

Existing image classification techniques highly rely on supervised models that are trained on large-scale datasets. Despite improved ontology engineering, such as ImageNet that includes 20K+ daily categories, the scale is far behind the requirement of generic image recognition. First of all, semantic concepts are complex and structured whereas the label space for most of current supervised learning consists of discrete and disjoint one-hot category vectors. The associations between classes are imposed to be neglected. Secondly, the dimension of label space is ever-growing. For example, on average, 1,000 new entries are added to Oxford Dictionaries On-line every. Consequently, for the scalability of conventional supervised learning is limited due to expensive acquisition of high-quality training images with annotations.

In the past decade, Zero-shod learning (ZSL) was proposed a potential solution which aim to transfer a learnt supervised model to unseen classes without acquiring new training

data at the test time. The essential problem is how to teach the machine what visual features will present in the test class using prior human knowledge. Therefore, the representation of human knowledge is required to maximumly bridge the visual-semantic gap. Most of existing approaches adopt visual attributes [15, 18, 26] so that a discrete class label can be embedded by a boolean representation, each dimension of which denotes whether an attribute present of absent. In this way, visual-attribute model from seen classes can be shared to unseen ones with using pre-defined attribute embeddings.

Although the generalisation to new classes can circumvent training image collection, constructing an attribute-based ontology is even more costly. As shown in Fig.1 (B), both seen and unseen classes need to be annotated by tens or hundreds of attributes. For example, the most popular benchmark, AwA, requires the annotator to give 85 attributes for each of 50 classes, let alone instance-level datasets, such as aPY and SUN which contain hundreds of thousands of manual annotations. Such restrictions severely prevent ZSL from being widely applied to many non-attribute scenarios. Furthermore, designing attributes is an ambiguous work since most of visual features are intangible. Constructing a large-scale ontology with attributes is thus time-consuming and error-prone. Recently, Demirel *et al.* [8] categories approaches without dedicated annotation efforts as *Unsupervised* Zero-shot Learning, which includes to use readily word embeddings [19], textual descriptions from the website [20], and hierarchical taxonomy information [2]. However, excluding tedious attribute annotation is at the cost of exhibiting a significant lower performance.

In this paper, we investigate how to spend the minimal annotation cost while still retain the high performance of that using attributes. Our first idea is inspired by an intuitive fact. To describe an unseen instance, the most straightforward way is to relate it to previously seen classes. Such expressions are called *Similes* [14] or *Classemes* [24] which explicitly compare two things by connecting words, *e.g. like, as, as, etc.* . However, most of existing approaches fail to quantify the simile between each pair of seen and unseen classes. And the annotation cost, as shown in Fig.1 (C), is not less than that using attributes. To this end, we propose a Fuzzy Interpolative Reasoning (FIR) algorithm that can infer full associations between seen and unseen classes using only a few similes. In our empirical study, we find only two similes are enough to outperform existing ZSL approaches using attributes. According to the common 40/10 seen/unseen split of AwA, the number of required labelling is reduced from 50×85 to only 10×2 that is only 0.47% of the original annotation work. Our main contribution is summarised as follows.

- We propose a simile-based ZSL learning framework that can significantly reduce the required annotation cost added to conventional supervised learning.
- The proposed FIR algorithm can effectively quantify similes and infer a reliable simile vector that can be used as improved semantic auxiliary information over conventional visual attributes.
- Despite low annotation cost, our approach outperforms state-of-art approaches including those using heavy-annotated attributes.

Related Work Simile refers to a part of speech which is proposed to describe complex visual features, e.g. facial similarity[14]. Another term is known as the Classemes [24] that can describe either objects similar to, or objects seen in conjunction with an unseen class, i.e. class-to-class-similarities. Existing methods often involve expensive class-to-class annotations [27], which is no difference to that of using attributes in terms of annotation cost.

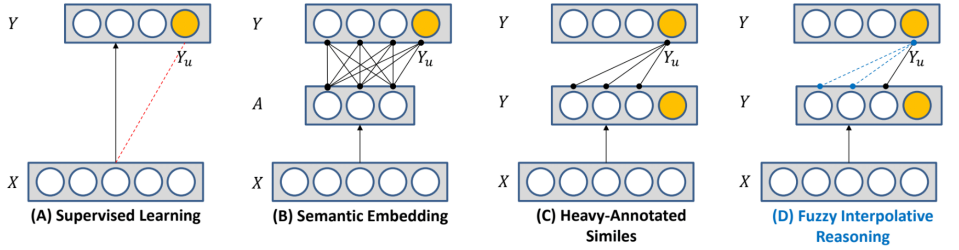

Figure 1: Zero-shot Learning framework comparison. Fuzzy Interpolative Reasoning leverages a few similes to infer the full associations of an unseen class to all of seen classes. Solid lines and arrows denote required annotations or associations.

Therefore, despite a large-scale simile-based ontology [16], such a stream of approaches have not gain much attention until some recent work [6, 17]. Contrary to these methods, our work utilise word embeddings as clues to find some initial similes to further minimise the human intervention. Due to the non-visual similarities of word embeddings, the human intervention is required to adjust the rank based on visual similarities and select a number of top similes. Despite light-weight annotations, the inferred representation through FIR can significantly boost existing ZSL methods by substituting their used attributes.

2 Methodology

2.1 Method Overview

1) Simile Vector Our framework is demonstrated in Fig.1 (D). Like most of conventional supervised classification using that in Fig.1 (A), given a set of seen classes $c \in 1,...,C$, a one-versus-all classifier ϕ_c is trained for each class. The classifier output is real-valued, *i.e.* $\phi_c(\mathbf{x}_i) > \phi_c(\mathbf{x}_j)$ implies that $\mathbf{x}_i$ is more similar to class c, where $\mathbf{x}$ denotes the visual feature vector of an image. Both training and test images then can be represented by $f_1(\mathbf{x}) = [\phi_1(\mathbf{x}),...,\phi_C(\mathbf{x})]$ which is referred as the *Simile Vector* in the simile space $\mathcal{V}$.

2) Fuzzy Interpolative Reasoning Since each dimension $v_i = \phi_i(\mathbf{x})$ denotes a similarity to a seen class, conventional approaches [27] exhaustively estimate all of seen classes to each unseen class and quantify the values by averaging the weights from a user questionnaire. However, such an paradigm suffers from subjective biases and requires heavy annotations, as shown in Fig.1 (C). We argue that a concise simile list is not only simple but also accurate for visual description. For example, it can be agreed that *leopard* looks like *bobcat* and *tiger* on a somehow close level. But it is difficult to decide either *rat* or *whale* is more dissimilar to *leopard*. Therefore, we model the problem as a fuzzy process where an unseen class can be represented by a membership function over top similes $[c_1,...,c_k]$ and later we introduce how to in turn convert the discrete similes to real values in the simile vector $[v_{c_1},...,v_{c_k}]$. The goal of FIR aims to modify each simile value in a complete vector $\mathbf{v}$, *i.e.* $f_{2_c}(v_{c_1},...,v_{c_k}) = v_c$.

Simile-based ZSL Suppose we have U unseen classes $u \in C+1,...,C+U$ that are disjoint from seen classes. Conventional supervised classifiers (Fig.1 (A)) cannot apply due to missing the link between images and these U class labels. This is known as the *Zero-shot Learning* problem. Now using the FIR inferred simile vector of each unseen classes ZSL prediction can be achieved by maximising a conventional compatibility score: $\mathbf{v}_1,...,\mathbf{v}_U$ $\arg\max_u f_3(f_1(\mathbf{x}),\mathbf{v}_u)$.

2.2 Fuzzy Interpolative Reasoning

A fuzzy membership function is defined as a measurable map:

$$f_2 := M(v) \rightarrow [0,1] \tag{1}$$

For example, given *leopard* is 0.9 similar to *bobcat*, the function can infer how much it is similar to *tiger* based on a rule base of observed membership, *i.e.* shared similarity between *bobcat* and *tiger*. Different from the probability theory that model predictions exclusively at a time, fuzzy process focus more on how much information in common at the same time. In this way, sparse values can likely get shared members to smooth the variance. In our case, we hope to maximise the tolerance to annotation errors, such as bad ranks or missing important similes due to the visual-semantic discrepancy.

Fuzzy Rule Base In this paper, we adopt the simplest Gaussian kernel as the membership function. Given a $\mathbf{x}$, K nearest neighbours $[\mathbf{x}_c^1, ..., \mathbf{x}_c^K]$ in class c are selected to estimated the similarity between $\mathbf{x}$ and class c:

$$\phi_c(\mathbf{x}) = \frac{1}{K}\sum_{i=1}^{K} \exp(-\frac{1}{2\sigma^2}\|\mathbf{x}_c^i - \mathbf{x}\|_2^2). \tag{2}$$

where $\sigma = 1$ without loss of generality. The similarity value is normalised to [0,1] using a sigmoid function. In this way, we can convert the whole training set $\mathcal{X}$ into the simile space $\mathcal{V}$ as the fuzzy rule base using $\mathcal{V} = [\phi_1(\mathcal{X}), ..., \phi_C(\mathcal{X})]$.

Simile Quantification For each unseen class, we have $[c_1, ..., c_k]$ similes. As argued earlier that it can be difficulty to give a specific value of the similarity and simile annotations can suffer from subjective variances. In this paper, we propose a novel empirical alternative that use the values in the rule base for initial quantification of discrete similes. Specifically, an unseen class has a simile of seen class $c_i \in [c_1, ..., c_k]$, we use the averaged self similarities of instances in c_i to make an approximation:

$$v'_{c_i} = \frac{1}{|c_i|}\sum_{v_i \in \mathcal{V}_{c_i}} v_i, \tag{3}$$

where $|.|$ is the cardinality of a class c_i; $\mathcal{V}_{c_i}$ denotes self-similarity values of all instances in class c_i: $\phi_{c_i}(\mathbf{x})$. In this way, we can in turn calculate the initial similarity values of the give similes $[c_1, ..., c_k] \rightarrow [v'_{c_1}, ..., v'_{c_k}]$. Next, we elaborate how to select proper observations in the rule base to complete the FIR algorithm: $f_{2_c}(v'_{c_1}, ..., v'_{c_k}) = v_c$ for C times to achieve a full simile vector $\mathbf{v} = [v_1, ..., v_c, ...v_C]$. Note that the initialised v's are also updated so as to further mitigate the annotation bias.

Fuzzy Rule Selection Like the membership estimation, it is not necessary to use all of the training instances as fuzzy rules. We only require rules that can make prominent effects to the conclusion. In this paper, we adopt a sparse manner to refine the fuzzy rule base. Firstly, we use the local region surrounding the observation instead of the global one, which is implemented by searching Q nearest neighbours of $[v'_{c_1}, ..., v'_{c_k}]$ in the c_k dimensional rule base. Afterwards, the *profile curvature* of the local region is constructed to represent the extent to which the local region deviates from being 'flat' or 'straight'. By viewing the pattern to be modelled as a geometry object, as shown in Fig.2, curvature values are used to estimate the prominence to the hidden pattern. As the 'flat' or 'straight' regions can be easily interpolated or approximated by its surroundings, only regions with higher curvature values

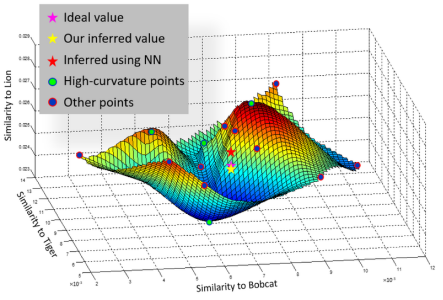

Figure 2: An demonstration of Spare fuzzy rule selection. The ideal value is computed by averaging the Leopard-Lion SMSs using real visual data. Using high-curvature points, the inferred value is more accurate than that using nearest neighbours. Note that the stars are in a vertical line and not on the surface. Some points are occluded.

need to be explicitly selected to construct fuzzy rule base. For simplicity, we introduce the following steps using a 2-D case as an example.

We employ the method in [23] to select the points with the steepest downward gradient for a given direction. Let the surface of the local region be denoted as $f(v_1, v_2)$, the gradient can be expressed as a 2-*D* vector field $\nabla f = (f_{v_1}, f_{v_2}, 0) = f_{v_1}^{(i)} + f_{v_2}^{(j)}$, where i and j are steps. The *slope* is defined as a scalar field:

$$S(v_1, v_2) = |\nabla f| = \sqrt{f_{v_1}^2 + f_{v_2}^2} \tag{4}$$

Using the S, the corresponding unit vector u is $u = (-\nabla f/S)$. For a given scalar field $F(v_1, v_2)$ the *directional derivative* D_u on the direction u and the overall profile curvature value K_v can be calculated:

$$D_u(F) = \nabla F \cdot u \tag{5}$$

$$K_v = -S^{-2}(f_{v_1}^2 f_{v_1 v_2} + 2 f_{v_1} f_{v_2} f_{v_1 v_2} + f_{v_2}^2 f_{v_2 v_2}), \tag{6}$$

where D_u compute the changing rate at F given a movement u; and K_v can be either positive or negative which corresponds to the convexity and the concavity respectively. For simplicity, we use *longitudinal profile curvature* that is a streamline passing through F(x,y). Firstly, we calculate eight directional derivatives for each point (clockwise from North to Northwest) which corresponds to the cardinal and inter-cardinal directions: $\{D_{u_1}, ..., D_{u_8}\}$. The point on each direction is interpolated using the *v4* function of Matlab toolbox *griddata* with parameter u as the density unit. The longitudinal profile curvature K_u can be calculated by comparing the pair of directional derivatives D_u and D'_u on the opposite directions ($D_u > D'_u$):

$$K_u = \frac{D_u - D'_u}{S^2} \tag{7}$$

Now the overall rule base $\mathcal{V}$ is refined into only top R points $[\mathbf{v}_1, ..., \mathbf{v}_R]$ with the highest K_u values. In this paper, we propose a novel multi-dimensional Gaussian membership function that can simultaneously accounts the R points, as shown in Fig.3 (R=5 here). For each of the r selected high-curvature points $\mathbf{v}_r$, the fuzzy set is constructed using its T nearest

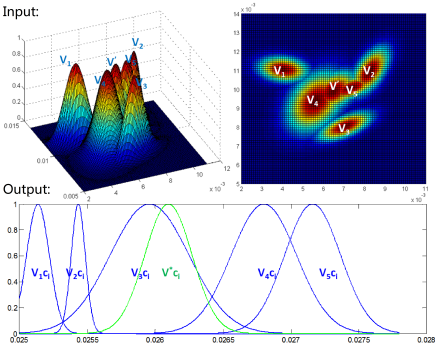

Figure 3: Fuzzy Rule Interpolation. $V_1 : V_5$ are refined high-curvature rules. $V_{1c_i} : V_{5c_i}$ are corresponding output fuzzy sets. $V^*_{c_i}$ is the final interpolation result.

neighbours $[\mathbf{v}_1, ..., \mathbf{v}_T]$ from the overall rule base. The membership function for each r point is:

$$M(\mathbf{v}_r) = \exp(-\frac{(\mathbf{v}_r - mu_r)^2}{2\Sigma_r^2}), \tag{8}$$

where $\mathbf{mu}_r$ and Σ_r are the mean and covariance of $[\mathbf{v}_1, ..., \mathbf{v}_T]$. Hereby, the representative value is updated from v_r to $Rep(\mathbf{v}_r) = \mathbf{mu}_r$ for further smoothing the data. Note that the membership value denotes the degree that a point belonging to the fuzzy set, where $M(\mathbf{mu}_r) = 1$.

Fuzzy Rule Interpolation As shown in Fig.3, the refined rule base is composed by R fuzzy rules. Given the initialised observation $\mathbf{v}' = [v'_{c_1}, ..., v'_{c_k}]$, we can construct its corresponding fuzzy set using eq.8. The new representative value is $Rep(\mathbf{v}) = mean(\mathbf{v}')$ instead of $\mathbf{v}'$. Our final step is to interpolate the real conclusion v_c using $\mathbf{v}$:

$$v_c = \frac{1}{R}\sum_{r=1}^{R} \lambda_r \mathbf{v}_r, \tag{9}$$

where λ_j is the interpolative ratio of the j^{th} fuzzy rule, which can be estimated by:

$$\lambda_r = \alpha_r \exp\left(-\frac{1}{2}(Rep(\mathbf{v}') - Rep(\mathbf{v}_r))^\top \Sigma_r^{-1} (Rep(\mathbf{v}') - Rep(\mathbf{v}_r))\right), \tag{10}$$

where $\alpha_j = \frac{1}{\sqrt{(2\pi)^2 \|\Sigma_j\|}}, j \in \{1, ..., r\}$. Finally, we can repeat the process from Eq.3 to Eq.10 for each seen class i to infer an SV class-level prototype: $v^* = [Rep(V_1^*), ..., Rep(V_C^*)]$.

Related Approaches Fuzzy logic theory provides an effective way to handle vague information that arises due to the lack of sharp boundaries or crisp values. With an inherent ability to analysis human natural language, the fuzzy logic theory has been widely applied to various intelligent systems [28]. Fuzzy interpolative reasoning is proposed to deal with the problem at where only sparse rule bases are available, that means, observations do not overlap with any rule antecedent values thus classical fuzzy inference methods have no rule to fire and cannot obtain any certain conclusions. It can be mainly categorised into two classes, the KH method [13] which is based on $\alpha-$cut of fuzzy set and interval algebra, and the HS methods [12, 22] which are based on representative values and analogical scaling and moving. Our

Table 1: Comparison to State-of-the-art Methods.

Annotation Type	Method	AwA		aPY	
		WE	SV	WE	SV
Unsupervised	DeViSE[11]	44.5	47.5	25.5	27.4
	ConSE[19]	46.1	48.2	22.0	27.8
	Text2Visual[9]	55.3	-	30.2	-
	SynC[5]	57.5	58.9	-	-
	ALE[1]	58.8	60.1	33.3	36.2
	LatEm[25]	62.9	63.2	-	-
	CAAP[2]	67.5	-	37.0	-
	Attri2Classname[8]	69.9	-	38.2	-
	Ours	-	**78.5**	-	**48.8**
Supervised		Attri	SV	Attri	SV
	DAP[15]	54.0	58.5	28.5	36.6
	ENS[21]	57.4		31.7	
	HAT[3]	63.1		38.3	
	ALE-attr[1]	66.7	70.1	-	-
	SSE-INT[29]	71.5	-	44.2	-
	SSE-ReLU[29]	76.3	-	46.2	48.9
	SynC-attr[5]	76.3	78.5	-	-
	SDL[30]	79.1	82.2	50.4	52.5
	Ours	80.5	**83.2**	51.2	**56.7**

WE: Word Embedding; SV: Simile Vector; Attri: Attribute Embedding.

unique contribution is to regard similes and visual similarities as fuzzy processes, for which we propose a new multi-variable multi-antecedent fuzzy interpolative reasoning algorithm that can accurately infer the similarity values.

2.3 Zero-shot Recognition Using SV

Using FIR, we can convert the light-weight sparse simile annotations into a full simile vector for each unseen class u $\mathbf{v}_u = [v_1, ..., v_c, ...v_C]$. During the test, an unseen image $\hat{x}$ is also converted into a simile vector $f_1(\hat{\mathbf{x}}) = \hat{v}$ using Eq.4. Without loss of generality, we simply adopt the simplest nearest neighbour classifier to predict the label:

$$f_3 := \arg\min_u \|\hat{\mathbf{v}} - \mathbf{v}_u\|_2^2 \tag{11}$$

3 Experiments

We first compare our approach to state-of-the-art results. Since simile-based ZSL has only a few previous work, our comparison involves published results under various settings, frameworks, and visual/semantic data. We discuss the characteristics of our method is details and try to understand how does each component contributes to the overall performance.

Our method is evaluated on AwA [15], and aPY [10] benchmarks. We follow standard 40/10 and 20/12 seen/unseen splits as that in [8] for the sake of fair comparison. We adopt the VGG-19 deep visual features released by [29]. Although the whole approach is non-parametric, there are four NN parameters: $[K, Q, R, T]$. The seen classes are divided into

Figure 4: Investigation of the characteristic of using similes.

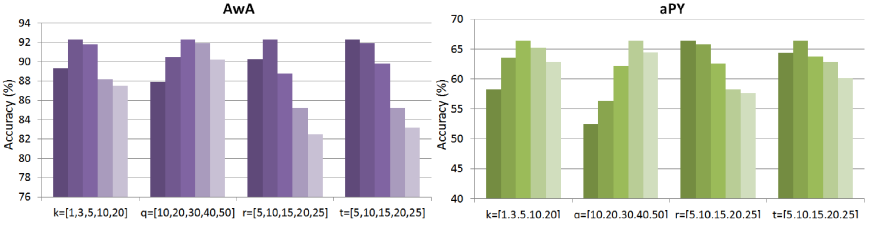

Figure 5: Performance *w.r.t.* nearest neighbour max-pooling parameters.

four folds. We use leave-one-fold-out cross-validation to choose the best parameters and fix them for all of the test.

One of the main concerns is how to achieve similes of unseen classes. In this paper, we defined three protocols: **1) Unsupervised Similes**: class similarities are purely estimated by their word embeddings and top c_k similes are fed to FIR; **2) Attribute-based Similes**: using conventional attributes as class embeddings, we can compute top c_k similes for FIR inputs; **3) Supervised Similes:** human intervention to the rank from protocol 1) to correct unsatisfied similes. The annotations use the judgements of a colleague who was unfamiliar with the details of this work.

3.1 Main Results

Baselines We summarise our main comparison in Table 1. It can be seen that most of the results of attribute-based approaches are better. The experimental results can be categories by two dimensions. Firstly, we compare our method to pure unsupervised approaches using word embedding-led similes. We then compare to conventional ZSL approaches using attributes to compute the similes. Furthermore, we provide light-weight human intervention to correct some simile errors due to semantic dominances from the above two embeddings. We focus on how much improvements can gain from the light-weight annotations. The averaged performance improvement is over 3.5%. Such a promising result indicate that our algorithm can provide an interesting interface for human-computer interaction to actively learn the parameters.

The other dimension is the comparison between conventional semantic models, i.e. WE and Attri, and the proposed SV. Also also implement some existing approaches with their released codes. Again, we observe significant performance gains by substituting WE and Attri by our inferred SV representation.

Our method steadily outperforms all of the above baselines. We ascribe our success to that encoding semantic similes by visual similarities between classes leads to little informa-

Table 2: Upper bound increase using SV as representation.

	AwA		aPY	
Feature	Deep	Low-level	Deep	Low-level
Raw	92.33	80.64	94.82	84.73
SV	96.83	91.88	97.42	95.62

tion loss comparing to attribute or word representations. Also, in contrast, the transductive setting is purely based on visual data distribution which may not be consistent with semantic distributions, whereas our similes are directly related to class labels.

3.2 Detailed Discussions

To understand the promising results, we discuss our approach from following aspects that are supported by extensive experiments.

Deep feature effect To understand the contribution of the SV representation, we separately study it using a supervised setting on seen classes. It can be seen that the supervised classification rates are remarkably increased, indicating the SV not only bridges the visual-semantic gap, but is a better visual representation as well. We verify our finds on conventional low-level features, e.g. a concatenation of PCA, PHOG, etc. to show its independence to deep features.

Simile Vector performance upper bound A recent novel evaluation metric is proposed in [7] to estimate the upper bound of the expected performance empirically. We absorb the same spirit and use the mean of real unseen visual data as prototypes. In Table 2, it can be seen that our SV can remarkably boost the upper bounds (roughly 4% for deep features and 10% for low-level features) than using raw features, which manifest our SV representation is not only interpretable but also more discriminative.

Reducing training samples In Fig. 4 (left), we gradually reduce the training size by randomly sampling a different number of images. The results are averaged by three repeats. It can be seen our method requires approximate 100 and 50 samples for AwA and aPY respectively to achieve reliable performance.

Simile error tolerance Semantic similes may be slightly inconsistent due to different interest points can be focused. Thus, in Fig. 4 (mid), we permute the top four similes and check their effects on the overall performance. It is shown that the first and second similes are more important than the third and fourth ones. aPY is more sensitive to simile errors because the number of seen classes (20) is much smaller than that of AwA (40). Hence, it can be difficult to give more than two similes to each unseen classes.

Increasing simile number A very interesting question is why we choose a pair of similes. We answer this question by the experiments in Fig. 4 (right). It is shown that using one simile is not satisfied. Our original assumption is that the performance would increase while more similes are given. To our surprise, using two, three, or four top similes does not make significant changes. Using five similes can harm the performance. The reason our FIR can regard correlated similes as one fuzzy set. Similes are not complementary to each other can lead to noise and redundancy. Thus, a pair is just enough.

NN parameters There are totally four NN processes in our approach. NN is a simple way for max-pooling on the feature level that can suppress noise and reduce redundancy. In Fig. 5, we show the effects of each parameter on the overall performance by fixing the other three. Generally, for MKE and high-curvature points, smaller k and r is better so as to pick

out high-quality observations. In contrast, higher q and t are better so that the rule base can contain sufficient fuzzy rules.

4 Conclusion

In this paper, we proposed an efficient simile-based ZSL framework that can recognise unseen classes with light-weight simile annotations. The proposed simile vector made the image representation more interpretable and discriminative. By regarding both input and output as fuzzy processes, our FIR significantly boosted the ZSL performance by accurately predicting the similarity value of each seen class in the simile vector using only discrete simile annotations. We achieved state-of-the-art results on both AwA and aPY. Our method significantly exceeded the performances of both supervised and unsupervised approaches. One of the future improvement could focus on how to apply similes on fine-grained tasks, which can be investigated in the future work.

Acknowledgements This work was supported in part by MRC Innovation Fellowship with ref MR/S003916/1.

References

[1] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for attribute-based classification. In *CVPR*, 2013.

[2] Zeynep Akata, Scott Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. Evaluation of output embeddings for fine-grained image classification. In *CVPR*, 2015.

[3] Ziad Al-Halah and Rainer Stiefelhagen. How to transfer? zero-shot object recognition via hierarchical transfer of semantic attributes. In *WACV*, 2015.

[4] Ziad Al-Halah, Makarand Tapaswi, and Rainer Stiefelhagen. Recovering the missing link: Predicting class-attribute associations for unsupervised zero-shot learning. In *CVPR*, 2016.

[5] Soravit Changpinyo, Wei-Lun Chao, Boqing Gong, and Fei Sha. Synthesized classifiers for zero-shot learning. In *CVPR*, 2016.

[6] Soravit Changpinyo, Wei-Lun Chao, and Fei Sha. Predicting visual exemplars of unseen classes for zero-shot learning. *ICCV*, 2017.

[7] Wei-Lun Chao, Soravit Changpinyo, Boqing Gong, and Fei Sha. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. *arXiv preprint arXiv:1605.04253*, 2016.

[8] Berkan Demirel, Ramazan Gokberk Cinbis, and Nazli Ikizler-Cinbis. Attributes2classname: A discriminative model for attribute-based unsupervised zero-shot learning. *ICCV*, 2017.

[9] Mohamed Elhoseiny, Babak Saleh, and Ahmed Elgammal. Write a classifier: Zero-shot learning using purely textual descriptions. In *CVPR*, 2013.

[10] Alireza Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. Describing objects by their attributes. In *CVPR*, 2009.

[11] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Tomas Mikolov, et al. Devise: A deep visual-semantic embedding model. In *NIPS*, 2013.

[12] Zhiheng Huang and Qiang Shen. Fuzzy interpolative reasoning via scale and move transformations. *IEEE Transactions on Fuzzy Systems*, 14(2):340–359, 2006.

[13] LászlóT Kóczy and Kaoru Hirota. Approximate reasoning by linear rule interpolation and general approximation. *International Journal of Approximate Reasoning*, 9(3): 197–225, 1993.

[14] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.

[15] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009.

[16] Li-Jia Li, Hao Su, Li Fei-Fei, and Eric P Xing. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *NIPS*, 2010.

[17] Yang Long and Ling Shao. Learning to recognise unseen classes by a few similes. In *ACMMM*, 2017.

[18] Yang Long, Li Liu, Fumin Shen, Ling Shao, and Xuelong Li. Zero-shot learning using synthesised unseen visual data with diffusion regularisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[19] Mohammad Norouzi, Tomas Mikolov, Samy Bengio, Yoram Singer, Jonathon Shlens, Andrea Frome, Greg S Corrado, and Jeffrey Dean. Zero-shot learning by convex combination of semantic embeddings. In *ICLR*, 2014.

[20] Marcus Rohrbach, Michael Stark, György Szarvas, Iryna Gurevych, and Bernt Schiele. What helps where–and why? semantic relatedness for knowledge transfer. In *CVPR*, 2010.

[21] Marcus Rohrbach, Michael Stark, and Bernt Schiele. Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In *CVPR*, 2011.

[22] Qiang Shen and Longzhi Yang. Generalisation of scale and move transformation-based fuzzy interpolation. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 15(3):288–298, 2011.

[23] Yao Tan, Jie Li, Martin Wonders, Fei Chao, Hubert P. H. Shum, and Longzhi Yang. Towards sparse rule base generation for fuzzy rule interpolation. In *FUZZ*, 2016.

[24] Lorenzo Torresani, Martin Szummer, and Andrew Fitzgibbon. Efficient object category recognition using classemes. In *ECCV*, 2010.

[25] Yongqin Xian, Zeynep Akata, Gaurav Sharma, Quynh Nguyen, Matthias Hein, and Bernt Schiele. Latent embeddings for zero-shot classification. In *CVPR*, 2016.

[26] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning-the good, the bad and the ugly. *CVPR*, 2017.

[27] Felix Yu, Liangliang Cao, Rogerio Feris, John Smith, and Shih-Fu Chang. Designing category-level attributes for discriminative visual recognition. In *CVPR*, 2013.

[28] Lotfi A Zadeh. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems*, 90(2):111–127, 1997.

[29] Ziming Zhang and Venkatesh Saligrama. Zero-shot learning via semantic similarity embedding. In *ICCV*, 2015.

[30] Ziming Zhang and Venkatesh Saligrama. Zero-shot learning via joint latent similarity embedding. In *CVPR*, 2016.