Zero-Shot Learning: Towards the Effortless Classification of Mystical Creatures.

1. The Problem.

Imagine a world in which you **have never heard of** or **seen a Pegasus**. However, you have seen many horses and birds. Now imagine someone describes to you what a Pegasus looks like. They tell you that it is a **flying horse**, with large, **plumed wings**.

You, as a human, using this semantic description, would easily be able to transfer your prior-knowledge about horses and birds to point out **which image is the Pegasus** within **Figure 1**.

This is the concept that **Zero-shot Learning** (ZSL) [3] is based upon and attempts to mirror. In ZSL, some form of side-information is used to **transfer knowledge** from the classes trained upon (the **"seen" classes**), to the **brand-new classes** introduced at **test time** (the "unseen" classes).

To achieve this task, **traditional** image-classification **techniques** would have to be **retrained on an entirely new dataset** of Pegasi. However, what if images of this new class are rare, e.g. in the case of a mythical creature? The harvesting of this new dataset could be **problematic and expensive**.

Figure 1. Image-classification challenge: can you transfer your prior-knowledge of horses and birds to identify the images of winged, flying horses?

2. Our Solution.

Most ZSL methods use **hand-annotated class-attributes**, such as "is white", "has wings" etc.. to transfer knowledge from seen to unseen classes. Unfortunately, these attributes are not always available, and annotating datasets can take up **1000s of hours of expert manual labor**.

In this project, we explore a less-commonly-used, but readily available, source of side-information, **Knowledge Graphs** (KG). KGs **link** together **everyday words and objects** to represent **human understanding** of language in a **machine-interpretable format**.

This allows both the **relationship inbetween classes** (such as Pegasus is similar to horse and bird) [2], and the **relationship between classes and any other concepts** (such as a Pegasus has wings and can fly, e.g. Figure 2) [4] **to be modelled** under a ZSL setting.

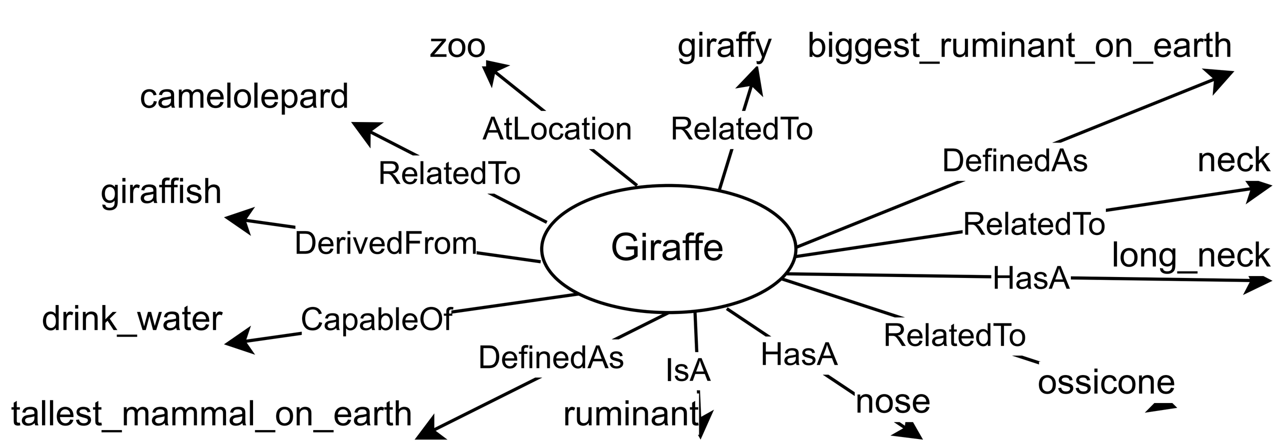
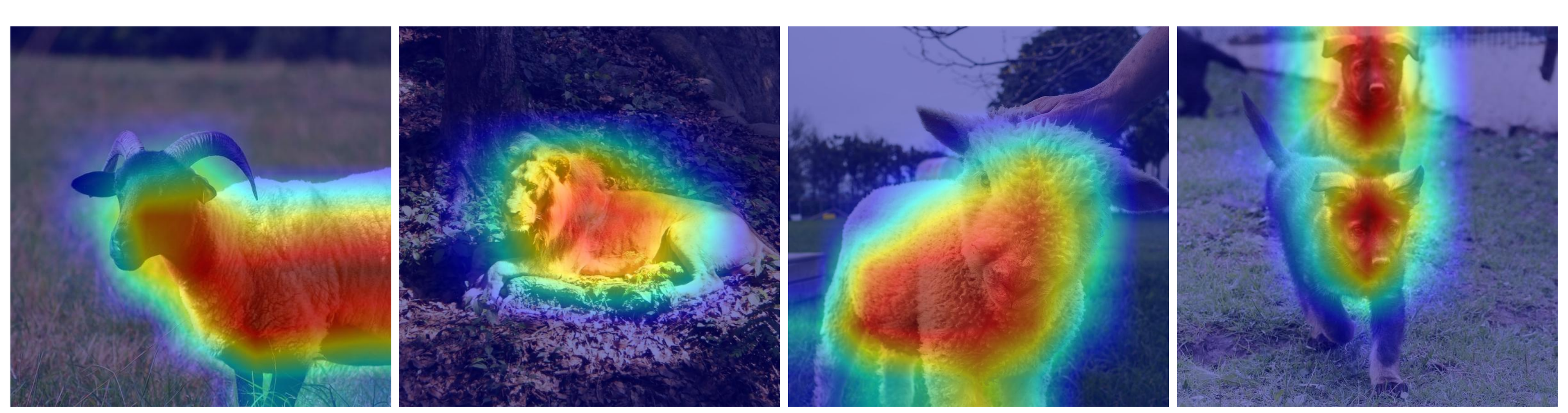


Figure 2. An extract of the rich, high-level information contained in the ConceptNet [6] KG for a giraffe.

Figure 3. Some correctly identified outputs of ZSL-KG+, with our website's GradCAM [5] heatmap enabled.



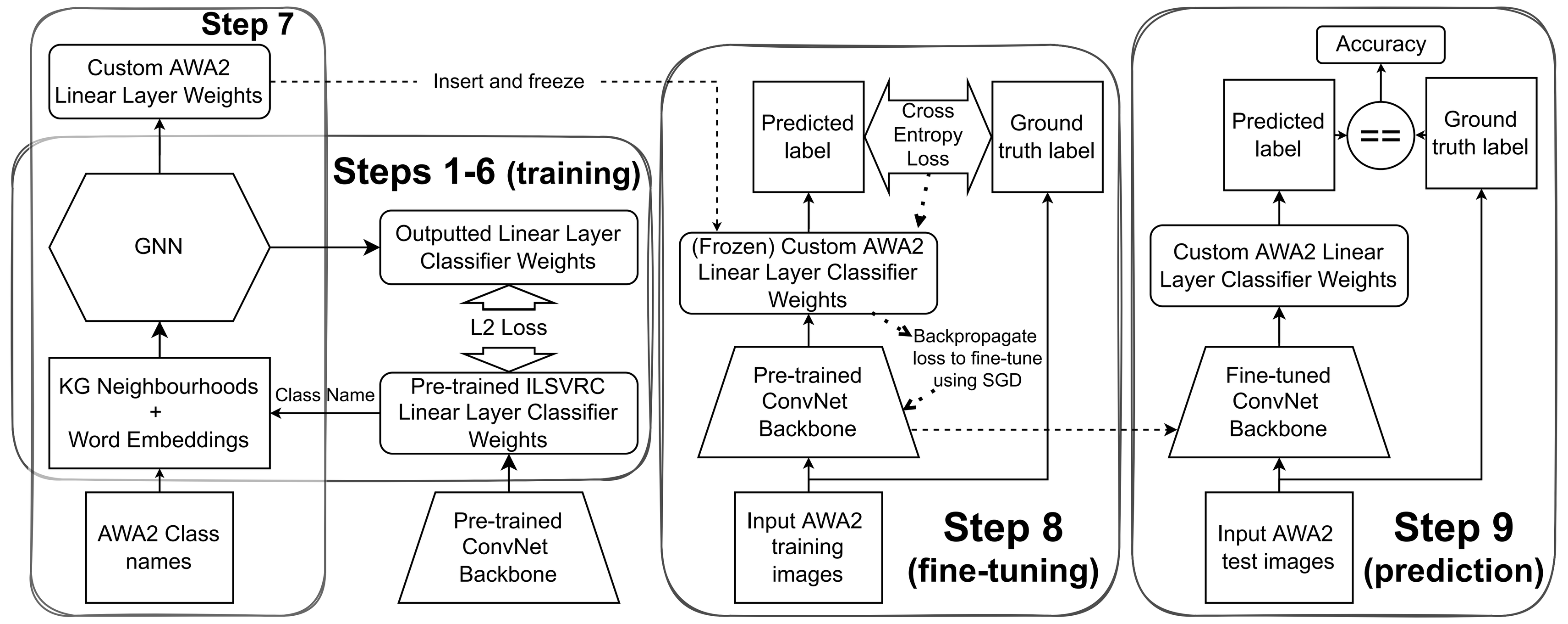


Figure 4. A diagrammatic representation of ZSL-KG+'s 9-step training process.

3. Our Results.

We propose ZSL-KG+, a improved framework based upon ZSL-KG [4], and evaluate on the Animals with Attributes 2 (AWA2) benchmark. ZSL-KG+ achieves:

- a new **state-of-the-art performance** of **82.1%** Top-1 accuracy in the ZSL setting;
- an **improvement of 4.0% accuracy** over the previous state-of-the-art model, see Table 1;
- a **58% reduction** in **model-parameters**;
- a **35% decrease** in **prediction-time**.

Overall, this makes ZSL-KG+ ideal for **low-budget time-critical real-world deployments**.

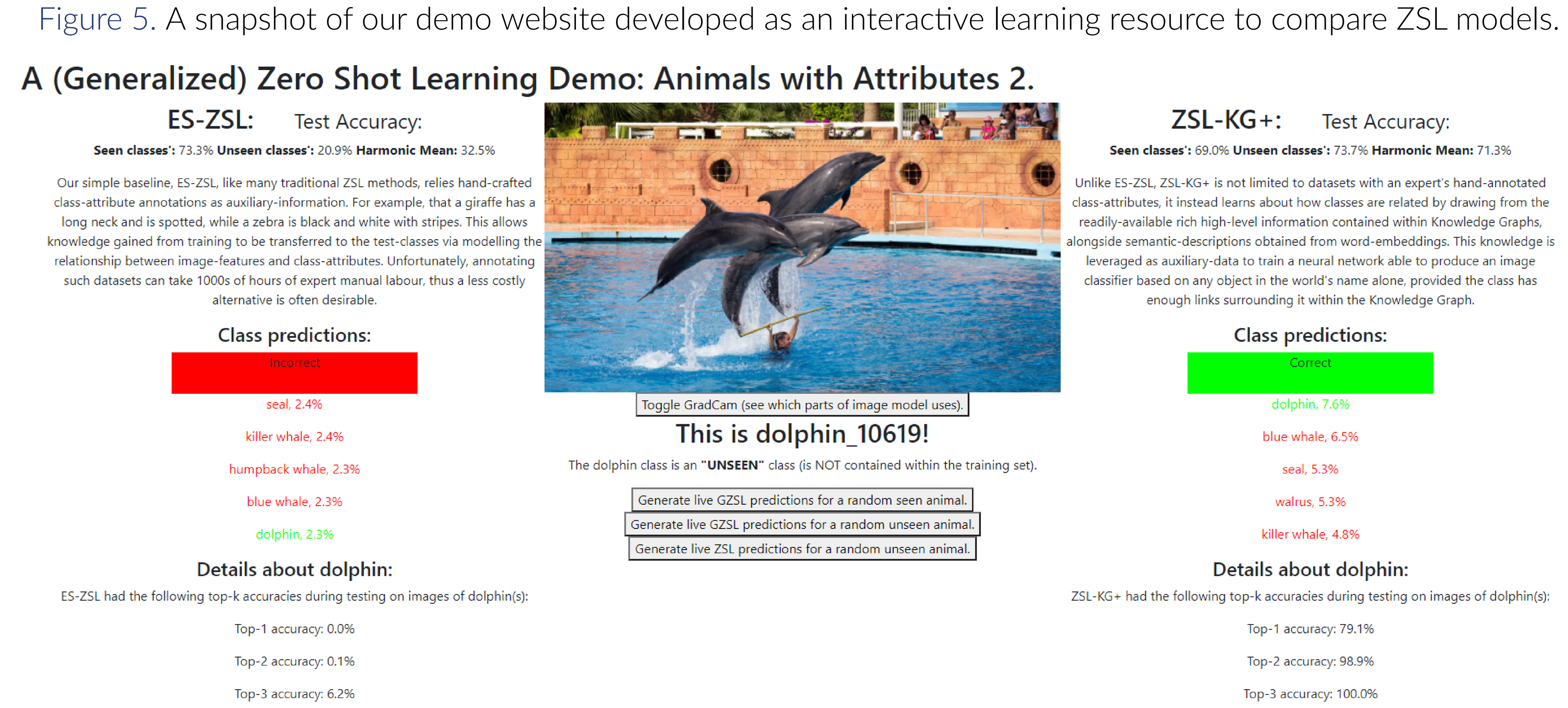
This is primarily achieved through **successfully integrating** the **EfficientNet** [7] **architecture** into ZSL for the first time. Furthermore, we **extensively ablate** the components and choices made within ZSL-KG+ and its predecessors in **ways not found in literature**.

Table 1. Comparison of ZSL-KG+ with best performing models from literature for ZSL and GZSL image-classification on the AWA2 dataset. We report the Top-1 accuracy ($T1$) for ZSL and harmonic mean (H) accuracy for GZSL.

	ZSML [8]	AGZSL [1]	DGP [2]	ZSL-KG [4]	ZSL-KG+
H	65.8	<u>76.8</u>	75.1	74.6	70.8
$T1$	77.5	76.4	77.1	78.1	<u>82.1</u>

We also delivered an **interactive** front-end **ZSL demo framework**, as presented in our website within Figure 5. This framework allows **live ZSL models** to be interacted with an analyzed using a variety of statistical tools, such as our **GradCAM heatmaps** displayed in Figure 3. This framework helps **improve the explainability** of the model, and is **novel** within literature.

Figure 5. A snapshot of our demo website developed as an interactive learning resource to compare ZSL models.

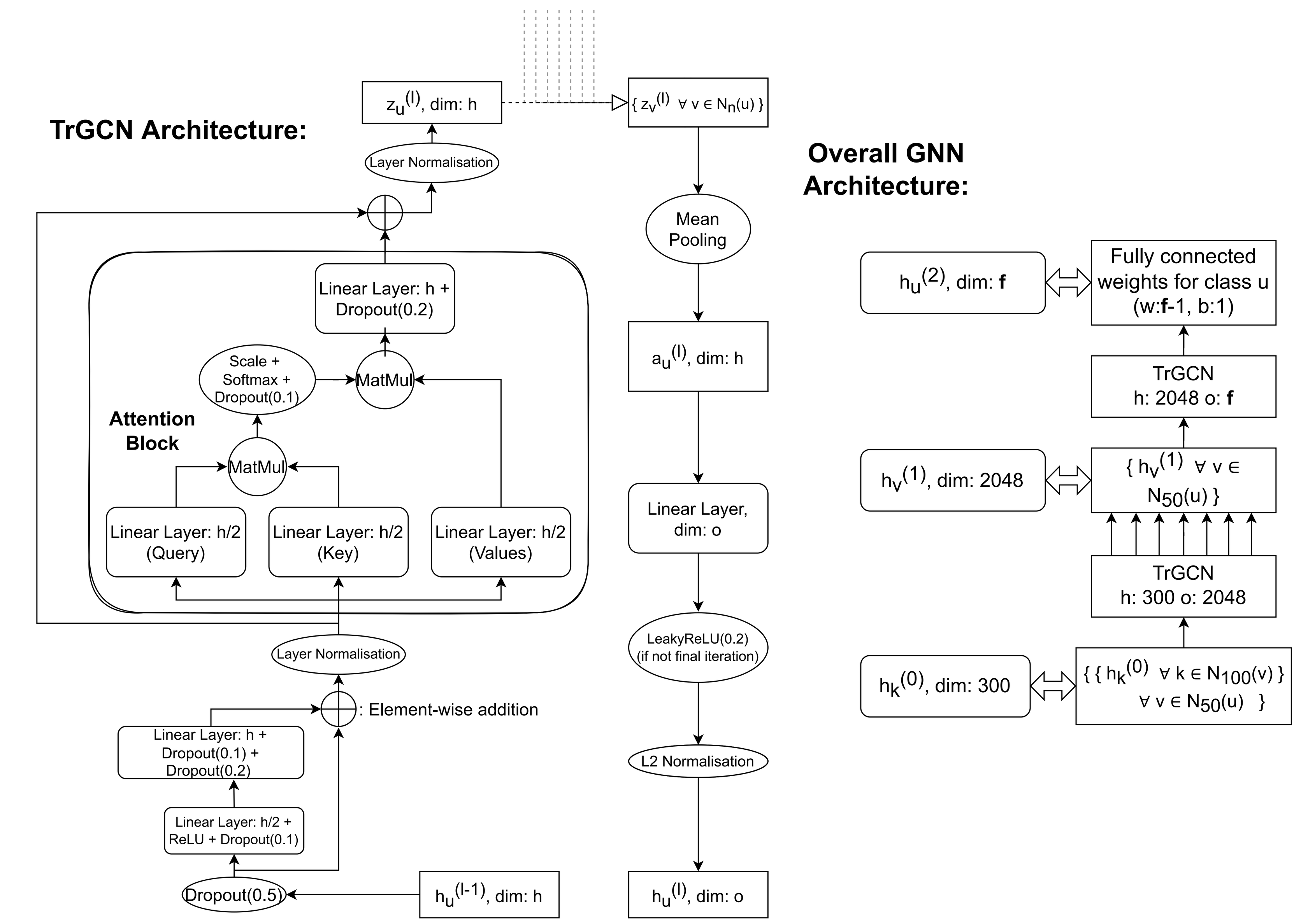


4. Our Methodology.

We leverage **semantic-relationships** from KGs, alongside **semantic-meanings from word-embeddings**, to train a **Graph Neural Network** (GNN; **Figure 6**) to produce the **weights of an image-classifier** for any novel class, provided only with its name, as per [4]:

1. **Random walks** are simulated to retrieve the class's "**local neighbourhood**" (**Figure 2**);
2. **For each neighbour**, additional random walks are simulated to retrieve **their neighbourhoods**;
3. Each original neighbour's neighbourhoods are run through a **transformer-based** TrGCN [4] aggregator to **summarise all their word-embeddings**;
4. The class's neighbour's aggregations are then run through the **second aggregator** to retrieve a final representation of the class, its **custom image-classifier weights**.

Figure 6. The architecture of our GNN, which aggregates graph-features in a 2-stage process, using the novel transformer-based TrGCN [4] aggregator module, to produce an image-classifier for an input class.



References

[1] Yu-Ying Chou et al. "Adaptive and generative zero-shot learning". *International Conference on Learning Representations*. 2020.

[2] Michael Kampffmeyer et al. "Rethinking knowledge graph propagation for zero-shot learning". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 11487–11496.

[3] Christoph H Lampert et al. "Learning to detect unseen object classes by between-class attribute transfer". *2009 IEEE conference on computer vision and pattern recognition*. IEEE. 2009, pp. 951–958.

[4] Nihal V. Nayak et al. "Zero-Shot Learning with Common Sense Knowledge Graphs". *arXiv:2006.10713* (2021).

[5] Ramprasaath R Selvaraju et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[6] Robyn Speer et al. "Conceptnet 5.5: An open multilingual graph of general knowledge". *Thirty-first AAAI conference on artificial intelligence*. 2017.

[7] Mingxing Tan et al. "Efficientnet: Rethinking model scaling for convolutional neural networks". *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.

[8] Vinay Kumar Verma et al. "Meta-learning for generalized zero-shot learning". *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 04. 2020, pp. 6062–6069.

Code: https://github.com/shadowbourne/ZSL-KG-Plus Student: Max Woolterton Supervisor: Yang Long Institute: Durham University Email: max.woolterton@durham.ac.uk