國立臺灣師範大學理學院
資訊工程學系

碩士論文

指導教授： 葉梅珍 博士

虛擬數據生成於零樣本學習

Generating virtual data for zero-shot learning

研究生： 李博衡 撰

中華民國 一一零 年 七 月

摘要

關鍵詞：零樣本學習

零樣本學習(zero shot learning)是指透過從已知類別的訓練資料中，將學習到的知識遷移到能辨識未知類別上。Li 的方法[22]透過語義嵌入和視覺特徵當作訓練資料，用深度學習模型將影像特徵轉變為語義分類器，並與語義嵌入做線性/非線性組合，藉此來找出與該視覺特徵相符的類別。

然而僅透過已知類別的資料當作訓練樣本，預測的成效可能有限，因為模型完全沒有對於未知類別的資訊。為了解決這個問題，本論文提出藉由隨機取樣的方式從已知類別的樣本中，產生出的新樣本，將其當作模擬未知類別的訓練樣本，也就是虛擬資料。讓模型在學習的時候，可以模擬未知類別的存在，使分類準確率得以提升。實驗於多個公開標準資料集驗證了所提出方法的可行性。

ABSTRACT

Keywords: zero-shot learning

Zero-shot learning refers to the knowledge transfer to recognize unseen classes, with a model learned from seen classes. Li's method [22] uses a deep learning method to transform visual embeddings into semantic classifiers, and performs linear / non-linear classification on semantic embeddings to predict the class.

Nevertheless, solely using data of seen classes may have a limited prediction performance because the model does not have any information of unseen classes. In order to solve this problem, we propose to generate virtual training data by randomly combining seen classes, which simulate unseen data. With such virtual data, the model can simulate the situation of recognizing unseen classes during learning, and therefore the classification accuracy can be improved. Experiments on several benchmark datasets have verified the effectiveness of the proposed method.

附表目錄

附圖目錄

目錄

第一章 簡介

在本章節中，我們將介紹零樣本學習的研究背景以及研究動機，闡述本研究中所觀察到零樣本學習在訓練過程中無法獲得未知類別資訊的問題，並以此提出透過隨機組合現有資料進行資料增強，作為解決該問題的方法。

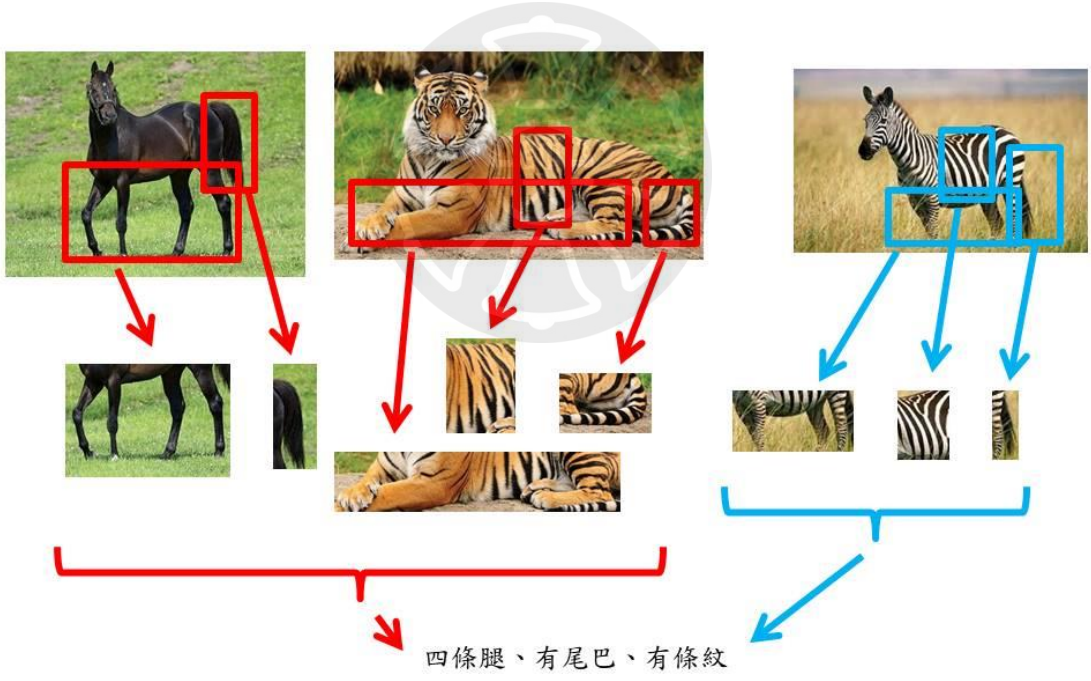
1.1 研究背景

在現實生活中，一般人可以用幾乎完全正確的準確率去分辨一張圖片中的物件是屬於哪一種類別，但是隨著需要辨識的圖片數陸續增多，人類的大腦就會開始無法負荷巨大的工作量，準確率可能也無法一路保持著相同的水準。因此，用深度學習的技術讓電腦學習如何辨識圖片開始逐漸成熟，但是，在訓練的過程中，需要準備大量已經標註過正確答案的圖片讓電腦可以正確的學習。然而礙於人力上的考量，每當要訓練電腦學習一種新的類別，就要花費大量成本去準備標注過的資料，不符合效益。所以，零樣本學習(zero-shot learning)，透過現有的訓練資源學習相關的先驗知識，實現對未知樣本進行準確的分類預測，這項技術開始逐漸被重視。

在零樣本學習的研究設置中，通常會加入類別的語義資訊當作輔助訊息，而這些語義資訊被稱為語義嵌入(semantic embedding)。在語義嵌入空間當中，概念比較接近且類似的類別，彼此的距離會比較靠近；反之，概念相異的類別間距離就會比較遠。透過這種特性，在解決零樣本學習的問題時，可以利用語義資訊

和視覺特徵建構出已知類別和未知類別的連結。

舉例來說，假如今天在辨識大量圖片時，其中有些圖片所屬的類別，沒出現在訓練樣本之中，比方說像是「斑馬」，在一般遇到沒學過相關的知識的情況下，不太可能進行準確的預測。但是，零樣本學習可以藉由「馬」、「老虎」這些動物的語義資訊進行組裝，加上「斑馬」的視覺特徵等一些輔助訊息，從中得知「斑馬」與「馬」和「老虎」有著共同的特徵，像是分別有四條腿、尾巴、有條紋等，藉此來把「馬」、「老虎」的相關知識遷移到名為「斑馬」這個新的類別，建立其與「馬」、「老虎」之間的關聯性，從而讓分類的準確率上升，示意圖如圖 1.1。



馬：https://number.bunshun.jp/articles/-/822693?page=3
老虎：https://www.chinatimes.com/hottopic/20210601002708-260809?chdtv
斑馬：https://daddypoppy.pixnet.net/blog/post/224137137-%E6%96%91%E9%A6%AC

圖 1.1 透過組合馬的四條腿和尾巴以及老虎的條紋，可以得到類似斑馬的特徵，以此建立彼此間的關聯性。

1.2 研究動機

零樣本學習(zero-shot learning)的最終目標為藉由學習已知類別的相關知識，將其遷移至未知類別，使在沒有看過未知類別的情況下，也能進行分類。在現有的方法中，透過深度學習模型把視覺特徵轉換成語義分類器，然後與語義嵌入做線性/非線性組合，以此來找出語義嵌入空間中與該視覺特徵相符的類別[8,22]。然而，上述方法在進行訓練時，使用的訓練樣本全都來自於已知類別的樣本，沒有額外考量到未知類別的樣本，造成未知類別的分類成效不佳。

為了解決這個問題，本論文提出一個增加訓練樣本的方法。透過從已知類別的訓練樣本中，隨機抽取類別以及選到的類別中的部份樣本，兩兩相互做線性組合，來產生虛擬的樣本，並以此當作未知類別的訓練樣本。藉由這個方法，可以讓模型在訓練的過程中，學習到大量各種已知類別的隨機組合，在進行測試的時候，可以提升分類的準確率；另外，增加已知類別以外的訓練樣本，可以讓模型模擬實際的測試情況，讓分類成效提升。

1.3 論文架構

本文將在第二章介紹近年的零樣本學習的方法，第三章描述模型的設計概念以及結構，並於第四章比較我們和其他方法於四種不同的公開資料集的實驗成效，最後於第五章提出結論。

第二章 相關工作

在本章節中，我們將描述語義嵌入的相關知識，並介紹近年來零樣本學習任務的相關工作。

2.1 語義嵌入

在處理零樣本學習的任務時，為了可以讓模型從已知類別學習到的知識，更順利的遷移到未知類別上，會加入語義嵌入這類的輔助訊息來幫忙，語義嵌入中包含每種類別對於各種屬性的相關描述。目前語義嵌入分為兩大類[5]：(1)工程語義嵌入(Engineered Semantic Spaces) 和(2)學習語義嵌入(Learned Semantic Spaces)。

工程語義嵌入是由人類手動進行標註，通常使用多個類別屬性(Attribute)做組合來當作語義嵌入，換句話說，每個維度人類都可以解釋出其涵義。舉例來說，「馬」具有「四條腿」、「無條紋」、「有尾巴」，其語義嵌入的形式可表示成[1,0,1]。

學習語義嵌入是透過機器學習的方式，讓模型自動產生類別的語義嵌入，通常無法解釋每個維度的個別涵義，例如 word2vec[6]。在標註成本方面，由於學習語義嵌入是自動產生的，比起工程語義嵌入，可以省下大量人力成本。

本研究中，我們選擇採用工程語義嵌入作為訓練模型的輔助訊息，因為在方法上，我們會透過線性組合現有樣本來生成虛擬樣本，為了讓虛擬樣本的可靠度上升，對於語義訊息的要求很高，因此人工標註的工程語義嵌入在優先度上，大

於學習語義嵌入。

2.2 零樣本學習

零樣本學習的目標為讓模型可以透過從現有的已知類別中，學習各種特徵表示，將學到的知識遷移至未知類別。根據實驗設置的嚴謹程度，可以區分為兩種：inductive 和 transductive 。前者在訓練模型的時候，訓練資料像是圖片或語義等，都只能來自已知的類別；而後者在訓練模型的時候，可以加入測試集中除了樣本對應標籤以外的樣本數據，藉此來學習測試集中的先驗知識。而在進行遷移學習的過程中，有些工作[10, 11]會在有未知類別的相關知識的前提下進行學習，讓學習效果變得更好，因此會採用 transductive 的設置。

除了用標註好的資料拿來訓練之外，有些工作[1, 2]選擇使用 GAN 等生成模型，自己生成未知類別的訓練樣本，來提高訓練資料的多樣性，達到使分類成果更好的效果；另外，也有些工作[3, 4]會搭配 attention 的技術，讓模型在訓練的過程中，專注在比較重要的區域性區塊，減少多餘資訊擾亂訓練成效的機會。例如，[3]在進行訓練的過程中，會補充來自影像特徵的自適應注意力，讓每個類的語義特徵可以得到拓展，使得不同類別之間(inter-class)和同一類別內之間(intra-class)的樣本有鑑別度；[4]則是根據屬性(attribute)的注意力機制，讓模型在學習時，把焦點放在與屬性比較相關的視覺特徵區塊上。

在評估零樣本學習的成效時，根據需要模型預測的內容，可以分為兩部分：

分別為 standard zero shot 和 generalized zero shot 。前者只需要模型預測輸入圖片屬於未知類別中的哪一項即可；而後者則需要模型預測輸入圖片是屬於全部類別中(包含已知類別和未知類別)的哪一項。後者在困難度方面，高於前者；而在實用程度上，後者也比較符合實際情況。

第三章 方法與步驟

在本章節中，我們首先將會定義零樣本學習的問題，其次再來介紹訓練資料和模型的設計理念以及架構。

3.1 問題定義

根據零樣本學習的設置，我們定義訓練資料總類別數 $C = C_s \cup C_u$，其中，C_s 為已知類別，C_u 為未知類別，且 $C_s \cap C_u = \emptyset$。訓練資料的格式方面，每一筆資料樣本可以表示成（x_i, y_i）的形式，x_i 表示影像視覺特徵；y_i 則是對應該視覺特徵的類別標籤，以 one-hot encoding 表示，即 $y_i = \{y_i^1, y_i^2 \dots y_i^{|c|}\} \in \{0,1\}$ 。而零樣本學習的最終目標，是想辦法將已知類別的相關知識，遷移至未知類別上，也就是學習一個分類器 $F: X \rightarrow Y$，使其在已知類別以及未知類別，可以完美的做出正確的分類。

在訓練的過程中，只能使用來自已知類別的訓練資料。而在測試方面，傳統的零樣本學習的測試過程中，測試樣本只需要預測未知類別即可；在廣義的零樣本學習當中，測試樣本需要預測的類別將擴大至所有的類別。

3.2 虛擬資料建置

根據[3, 7]，在訓練模型的過程中，輸入的訓練資料除了來自資料集本身提供的可見類樣本、語義嵌入之外，我們會額外從可見類中隨機抽選出 m 個類，

並分別從剩餘沒被選到的可見類中，選擇餘弦相似度最相近的做搭配，最終得到 $2m$ 個(共 m 對)不重複的類別。而每個類別隨機取 n 個樣本，用選到的類別的樣本(x_i與x_j)和語義嵌入(a_i與a_j)，兩兩互相做線性的組合(如公式 1 與公式 2)，最終得到 m 對新的類別，稱為虛擬資料（virtual data)：

$$a_v = \lambda a_i + (1-\lambda) a_j \quad (1)，$$

$$x_v = \lambda x_i + (1-\lambda) x_j \quad (2)$$

$\lambda \in [0,1]$來自 beta 分佈[7]，a_v 為虛擬類別語義嵌入，x_v 為虛擬類別視覺特徵。

3.3 模型架構

模型架構方面，根據[8, 22]，我們把每一張圖像轉換成獨立的語義分類器，藉由加入語義嵌入訊息以及損失函數，來達成分類的目的。而語義分類器的做法，是將預先訓練好的卷積神經網路(CNN)當作特徵抽取器，把原始影像透過其轉換成影像視覺特徵 X，接著經過兩層全連接層，也就是 $f(\cdot)$，得到語義分類器參數，即：

$$f(X) = W = (M, b) \quad (3)$$

而語義分類器可以被定義為：

$$F(a, W) = \arg\max (D(a, W)) \quad (4)$$

，其中，a 為語義嵌入特徵，D 為分類器方程式，W 為分類器參數。

分類器方程式方面，可以分為線性以及非線性版本。在線性分類器中，透過 $f(\cdot)$輸出的兩個分類器參數(M, b)，得到 $M(a)+b$ ($M \in R^{1xd}$，$b \in R^1$)；而非線性分類器，藉由引入超參數 k，$f(\cdot)$會輸出的分類器參數可分成四個(M_1，M_2，b_1，b_2)，得到$M_2(\tanh(M_1(a) + b_1)) + b_2$（$M_1 \in R^{kxd}$，$M_2 \in R^{1xk}$，$b_1 \in R^k$，$b_2 \in R^1$)。在進行預測時，輸入語義嵌入向量$a \in R^{dxn}$給線性/非線性分類器，再通過 Softmax 可以得到該影像對應到各個類別的機率值，從中選擇機率值較大的類別當作預測的結果，其中，d 為語義嵌入維度，n 為類別個數。

架構圖如下所示：

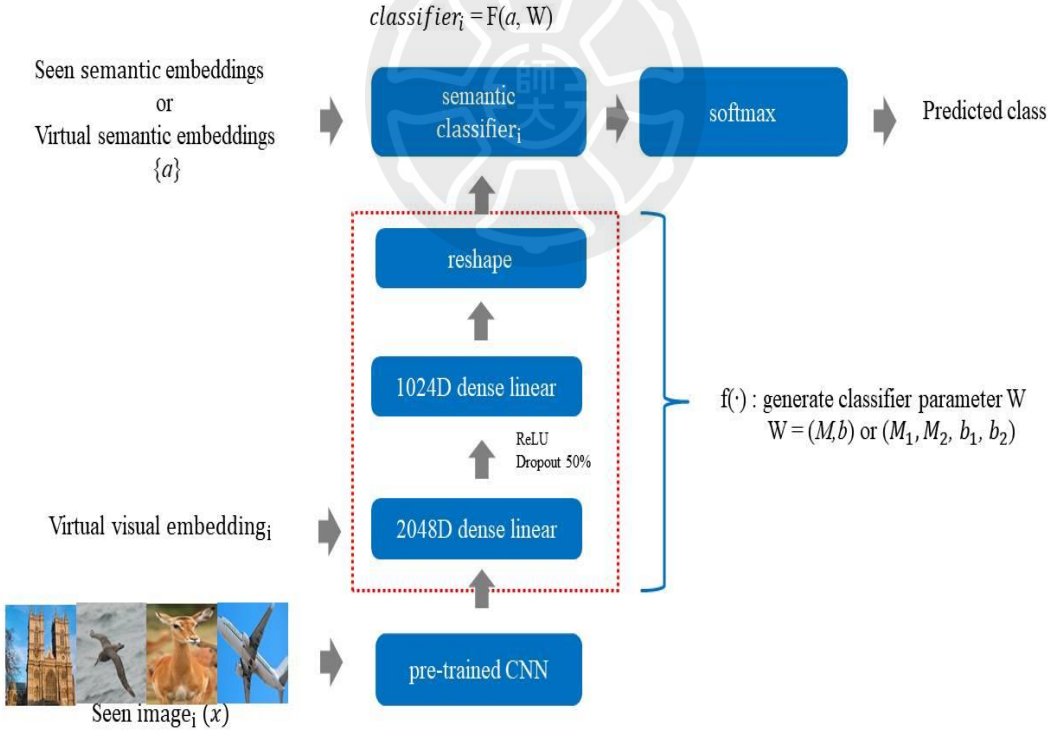


圖 3.1 模型架構圖

在模型訓練階段，每一個 epoch 會分別輸入已知類別的訓練資料以及重新生成的虛擬訓練資料，其中，虛擬訓練資料的用處在於模擬未知類別的測試資料，讓模型在訓練的過程中，模擬真正在測試階段時所遇到的設置。

而損失函數的部分，我們採用 cross entropy：

$$-\sum_{i=1}^{N}\sum_{j=1}^{|C|}(y_i^j \ln(p_i^j) + (1-y_i^j)\ln(1-\ln(p_i^j))) \quad (5)$$

其中，N 為樣本數，C 為類別數，y 為 ground truth，p 為模型預測的機率值。

第四章 實驗

在本章節中，我們將介紹資料集以及零樣本學習的實作細節，並說明如何進行評估所提出的方法。

4.1 資料集

資料集	語義嵌入維度	類別個數		樣本個數			
		已知類別	未知類別	訓練及驗證樣本	已知類別測試樣本	未知類別測試樣本	樣本總數
CUB[23]	312	150	50	7,057	1,764	2,967	11,788
SUN[24]	102	645	72	10,320	2,580	1,440	14,340
AWA2[9]	85	40	10	23,527	5,882	7,913	37,322
APY[25]	64	20	12	5,932	1,483	7,924	15,339

表 4.1 四個標準資料集的資料

如表 4.1 所示，資料集方面，根據[9]的設置，我們採用四個公開資料集作為零樣本學習的訓練資料以及測試資料，分別為 Caltech-UCSD-Birds 200-2011(CUB) [23]、SUN(SUN) [24]、Animals with Attributes 2(AWA2) [9]、Attribute Pascal and Yahoo(APY) [25]。CUB 資料集為鳥類的中等規模大小資料集，專門用於進行細緻化分類，總共包含 200 種鳥類，其中 150 種鳥類用於訓練，剩

餘 50 種鳥類用於測試；SUN 資料集為用於場景辨識的中等規模大小資料集，也是專門用於進行細緻化分類，總共包含 717 種室內外場景，其中 645 種場景用於訓練，剩餘 72 種場景用於測試；AWA2 資料集為動物的資料集，其中 40 種動物用於訓練，剩餘 10 種動物用於測試；APY 為小規模資料集，總共有 32 種類別，其中 20 種類別用於訓練，剩餘 12 種類別用於測試。另外，我們使用類別的屬性資料作為語義嵌入，來輔助零樣本學習任務。

4.2 實作細節

在實作上，我們用來抽取視覺特徵的 CNN 模型，採用殘差網路 ResNet-101，參數方面是使用在大規模單標籤影像資料集 ImageNet 上面預先訓練好的權重參數。我們將最後一層的全連接層移除後，加入兩層全連接層，也就是 f(・)，來形成語義分類器，而在進行訓練時，會凍結抽取特徵的參數部分，只會訓練到 f(・)的參數而已，其中，在非線性語義分類器的設置下，超參數 k 設為 30。優化器方面，我們使用 adam 並設定學習率為 1e-4。生成虛擬資料方面，隨機產生的 λ 的 beta distribution 的α以及β參數皆設為 0.4，另外，我們會根據資料集以及語義分類器設置的不同，擬定產生不同數量的虛擬類別及樣本，詳見表 4.2、表 4.3，其中，非線性版本因為模型較為複雜，樣本數設置的比線性版本少，避免擾亂訓練；而 AWA2 因為每個可見類的平均樣本數過高，同樣為了避免擾亂訓練，所以在線性版本設置的樣本數偏小。實驗是運行在配備 Intel Core i5-7500 3.4-GHz，

NVIDIA GeForceGTX 1080Ti 和 24 GB RAM 的機器上。

	SUN	CUB	AWA2	APY
class	322	75	20	10
Samples	80	80	20	80

表 4.2 線性語義分類器下虛擬資料的設置

	SUN	CUB	AWA2	APY
class	322	75	20	10
Samples	20	20	20	20

表 4.3 非線性語義分類器下虛擬資料的設置

4.3 評估方式

為了評估我們的方法對於分類已知類別以及未知類別的可行性，我們採用零樣本學習(ZSL)和廣義零樣本學習(GZSL)設置來進行評估。對於 ZSL 設置，測試影像的類別屬於未知類別(X_u)，預測類別僅需包含未知類別(C_u)；而對於 GZSL 設置，測試影像的類別屬於已知類別或未知類別(X_s or X_u)，預測類別則皆包含已知和未知類別(C)。

評估指標方面，為了數值化零樣本學習的學習成果，通常會採用 top-1 準確率，其數值代表預測結果排序第一名的類別和正確結果相符的比例。但是透過上

述方式來衡量準確率，有可能會忽略那些類別樣本數比較少的測試資料對整體正確率帶來的影響，因此，我們採用平均類別的 top-1 準確率(average per-class top-1 accuracy)做為評估指標，公式如下：

$$\text{acc} = \frac{1}{|C|}\sum_{i=1}^{|C|}\frac{\text{類別 i 答對的樣本數}}{\text{類別 i 的樣本總數}} \quad (6)$$

而 GZSL 設置中，我們希望已知類別以及未知類別的樣本都可以展現出優秀的預測能力，因此，我們採用調和平均數 H(harmonic mean)來評估廣義零樣本學習的成效，公式如下：

$$\text{H} = \frac{2 * acc_{X_s \to C} * acc_{X_u \to C}}{acc_{X_s \to C} + acc_{X_u \to C}} \quad (7)$$

4.4 實驗一：數據比較

方法	SUN	CUB	AWA2	APY
PQZSL[16]	-	-	-	27.9
CONSE[17]	38.8	34.3	44.5	26.9
DEVISE[18]	56.5	52	59.7	39.8
Relation Net[19]	-	55.6	64.2	-
SP-AEN[13]	59.2	55.4	58.5	24.1
PSR[14]	61.4	56	63.8	38.4
CDL[15]	63.6	54.5	-	43
IGSC(linear)[22]	55.4	51.9	58.2	36.5
IGSC(nonlinear)[22]	58.3	56.9	62.1	35.2
LFGAA+Hibrid[20]	62	67.6	68.1	-
AGZSL[3]	63.3	57.2	73.8	41
Ours(linear)	58.89	56.26	65.88	37.04
Ours(nonlinear)	57.08	56.46	62.98	35.76

表 4.4 傳統零樣本學習在 SUN、CUB、AWA2 和 APY 的實驗結果，以 top-1 精準度百分比表示

方法	SUN			CUB			AWA2			APY		
	u	s	H	u	s	H	u	s	H	u	s	H
SP-AEN[13]	24.9	38.6	30.3	34.7	70.6	46.6	23.3	90.9	37.1	13.7	63.4	22.6
PSR[14]	20.8	37.2	26.7	24.6	54.3	33.9	20.7	73.8	32.3	13.5	51.4	21.4
CDL[15]	21.5	34.7	26.5	23.5	55.2	32.9	-	-	-	19.8	48.6	27.1
CONSE[17]	6.8	39.9	11.6	1.6	72.2	3.1	0.5	90.6	1	0	91.2	0
DEVISE[18]	16.9	27.4	20.9	23.8	53	32.8	17.1	74.7	27.8	4.9	76.9	9.2
PQZSL[16]	35.1	35.3	35.2	53.2	51.4	46.9	-	-	-	64.1	64.1	38.8
Relation Net[19]	-	-	-	38.1	61.1	47	30	93.4	45.3	-	-	-
LFGAA+Hibrid[20]	18.5	40	25.3	36.2	80.9	50	27	93.4	41.9	-	-	-
IGSC[22]	39.4	31.3	34.9	40.8	60.2	48.7	25.7	83.6	39.3	23.1	58.9	33.2
LsrGAN[21]	44.8	37.7	40.9	48.1	59.1	53.0	54.6	74.6	63.0	-	-	-
AGZSL[3]	29.9	40.2	34.3	41.4	49.7	45.2	65.1	78.9	71.3	35.1	65.5	45.7
Ours (線性)	20.7	32.3	25.2	27.6	62.2	38.2	17.9	90.6	29.8	11.7	68.9	19.9
Ours (線性+CS)	39.5	25.6	31.1	46.9	53.5	50.0	43.0	73.1	54.2	30.1	46.0	36.4
Ours (非線性)	21.0	31.9	25.3	26.0	64.1	37.0	21.4	87.0	34.3	17.0	62.1	26.6
Ours (非線性+CS)	34.1	29.6	31.7	43.1	51.1	46.8	30.4	85.7	44.9	28.2	51.4	36.4

表 4.5 廣義零樣本學習在 SUN、CUB、AWA2 和 APY 的實驗結果，以 top-1 精準度百分比表示

表 4.4 和表 4.5 分別列出在傳統零樣本學習以及廣義零樣本學習的實驗結果，並與其他的零樣本方法做比較。我們的方法是基於 IGSC[22]做改善，可以發現加入虛擬資料後，在 SUN、CUB、AWA2 和 APY 資料集上準確率有所提升；而與最近的方法相比，我們與有使用 GAN 的方法像是 LsrGAN[21]和注意力機制的方法像是 AGZSL[3]，在準確率方面有所落差。根據[12]，在廣義零樣本學習的

設置下，因為訓練資料都是來自於已知類別的緣故，模型預測已知類別的機率值分數通常會大於未知類別的機率值分數，所以會加入一個校準因子來扣掉已知類別的分數，藉此來平衡已知類別和未知類別的機率值得分，在 SUN、CUB、AWA2、APY 的校準因子分別設為 0.2、0.5、0.3、0.5，校準(CS)公式如下：

$$P_i = (P_i - \gamma 1[i \in C_s]) \quad (8)$$

其中，$1[i \in C_s] \in \{0,1\}$ 用來判斷該類別是否屬於可見類，γ為校準因子。

4.5 實驗二：Ablation Study

		GZSL			ZSL
		u	s	H	
線性	可見類資料	19.1	24.6	21.5	55.4
	虛擬資料	20.56	30.47	24.55	57.57
	虛擬資料加可見類資料	20.69	32.33	25.23	58.89
非線性	可見類資料	22.5	36.1	27.7	58.3
	虛擬資料	20.35	31.24	24.64	57.15
	虛擬資料加可見類資料	20.97	31.9	25.31	57.08

表 4.6 在 SUN 資料集，比較只用可見類資料、虛擬資料以及虛擬資料加可見類訓練資料的效果

		GZSL			ZSL
		u	s	H	
線性	可見類資料	16.5	67.4	26.4	58.2
	虛擬資料	20.29	72.92	31.74	61.4
	虛擬資料加可見類資料	17.86	90.58	29.83	65.88
非線性	可見類資料	19.8	84.9	32.1	62.1
	虛擬資料	17.17	63.03	26.99	54.24
	虛擬資料加可見類資料	21.39	87.03	34.34	62.98

表 4.7 在 AWA2 資料集，比較只用可見類資料、虛擬資料以及虛擬資料加可見類訓練資料的效果

表 4.6 和表 4.7 分別列出當資料集為 SUN 和 AWA2 時，訓練資料只使用可見類資料、虛擬資料和使用虛擬資料加上可見類資料的情況下，對於分類結果的影響。

透過觀察表 4.6 和表 4.7，我們可以發現如果訓練資料只使用可見類資料或是生成的虛擬資料，在分類成效上，普遍輸給使用虛擬資料加上可見類訓練資料當訓練樣本的情況，因為可見類資料加入虛擬資料進行資料增強，可以有效控制在學習遷移知識時模型的複雜度，不會讓模型過度偏向學過的類別，讓分類效果提升；其中，虛擬樣本是組裝出來的，與實際的 unseen class 有所差異，非線性分類器因為模型較複雜，容易受虛擬樣本影響，效果沒有線性分類器來得好，

而 SUN 屬於 fine-grained 資料集，加入虛擬樣本後的效果上 AWA2 較好。因此，在訓練階段的訓練樣本使用虛擬資料加上可見類訓練資料會比較好。

		GZSL			ZSL
		u	s	H	
線性	隨機配對組裝	19.38	29.81	23.48	56.94
	取最相似組裝	20.69	32.33	25.23	58.89
非線性	隨機配對組裝	21.32	34.03	26.22	57.57
	取最相似組裝	20.97	31.9	25.31	57.08

表 4.8 在 SUN 資料集，比較虛擬資料生成時隨機配對組裝以及有取最相似組裝的效果

		GZSL			ZSL
		u	s	H	
線性	隨機配對組裝	16.28	70.52	26.46	65.16
	取最相似組裝	17.86	90.58	29.83	65.88
非線性	隨機配對組裝	20.61	89.26	33.49	61.51
	取最相似組裝	21.39	87.03	34.34	62.98

表 4.9 在 AWA2 資料集，比較虛擬資料生成時隨機配對組裝以及有取最相似組裝的效果

表 4.8 和表 4.9 分別列出當資料集為 SUN 和 AWA2 時，在生成虛擬資料的階段，隨機配對進行組裝和有取最相似進行組裝的情況，對於分類結果的影響。

透過觀察表 4.8 和表 4.9，我們可以發現在生成虛擬資料的階段，隨機配對進行組裝，在分類成效上，普遍輸給有取最相似進行組裝的情況，因為如果只是單單進行隨機組合，有些組合在邏輯上明顯不可能出現，若是讓這些組合加入訓練資料之中，會打亂訓練成效；而在組裝時，取相似度較高的特徵進行組裝，則可以有效避免上述提到的問題，讓測試的分類效果提升。因此，在生成虛擬資料時，根據相似度取最相似的進行組裝，在分類效果上會比較好。

		SUN			CUB			AWA2			APY		
		u	s	H	u	s	H	u	s	H	u	s	H
線性	加入 unseen 資訊	19.1	26.24	22.11	26.35	57.84	36.2	18.01	90.27	30.03	11.79	67.66	20.07
	無加入 unseen 資訊	20.69	32.33	25.23	27.6	62.22	38.24	17.86	90.58	29.83	11.65	68.86	19.94
非線性	加入 unseen 資訊	19.93	32.29	24.65	25.19	64.16	36.17	21.29	88.81	34.35	14.82	66.21	24.22
	無加入 unseen 資訊	20.97	31.9	25.31	26.02	64.12	37.02	21.39	87.03	34.34	16.96	62.09	26.64

表 4.10 在廣義零樣本學習設置下，比較虛擬樣本組裝時加入 unseen 資訊的效果

		SUN	CUB	AWA2	APY
線性	加入 Unseen 資訊	55	53.63	65.41	36.18
	無加入 Unseen 資訊	58.89	56.26	65.88	37.04
非線性	加入 Unseen 資訊	56.18	55.05	62.66	36.77
	無加入 Unseen 資訊	57.08	56.46	62.98	35.76

表 4.11 在傳統零樣本學習設置下，比較虛擬樣本組裝時加入 unseen 資訊的效果

表 4.10 和表 4.11 分別列出在廣義零樣本學習和傳統零樣本學習設置下，在生成虛擬資料的階段，加入 unseen class 的資訊，從 seen class 中選擇與其高相似的類別產生配對以及 seen class 自己做高相似配對，對於分類結果的影響。

透過觀察表 4.10 和表 4.11，我們可以發現在生成虛擬樣本時，加入 unseen class 的資訊，從 seen class 中選擇與其高相似的類別產生配對，在分類成效上，普遍輸給 seen class 自己做高相似配對的情況，因為使用前者的方法，產生的虛擬樣本類別個數(等同於未知類別的個數)小於後者的方法，在分類成效上會比較不好。因此，在生成虛擬資料時，seen class 自己做高相似配對，在分類效果上會比較好。

第五章 結論

在本論文中，對於零樣本學習任務，有別於先前的工作，在訓練的過程中只使用已知類別的資料，我們針對已知類別的訓練樣本作隨機的線性組合，額外產生出虛擬數據當作訓練樣本，藉此模擬測試時未知類別的數據，達到資料增強的效果。

我們所提出的方法與 IGSC[22]相比，發現加入虛擬資料後，在準確率上有所提升。而未來研究方向，我們預期朝向研究更適合的 meta learning 方法可以跟虛擬資料做搭配，讓分類準確率可以更上一層樓。

參考著作

[1] ZongYan Han, Zhenyong Fu, Jian Yang,“Learning the Redundancy-free Feature for Generalized Zero-Shot Object Recognition”,in *CVPR*,2020.

[2] Jiamin Wu, Tianzhu Zhang, Zheng-Jun Zha, Jiebo Luo, YongDong Zhang, Feng Wu,“Self-supervised Domain-aware Generative Network for Generalized Zero-shot Learning”,in *CVPR*,2020.

[3] Yu-Ying Chou, Hsuan-Tien Lin, Tyng-Luh Liu,“ADAPTIVE AND GENERATIVE ZERO-SHOT LEARNING”, in *ICLR*,2021.

[4] Dat Huynh, Ehsan Elhamifar,“Fine-Grained Generalized Zero-Shot Learning via Dense Attribute-Based Attention”, in *CVPR*,2020.

[5] WEI WANG, VINCENT W.ZHENG, HAN YU, CHUNYAN MIAO,“A Survey of Zero-Shot Learning: Settings, Methods, and Applications”,in *ACM Transactions on Intelligent Systems and Technology(TIST)*,2019.

[6] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean,“Efficient Estimation of Word Representations in Vector Space”,in *ICLR*,2013.

[7] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, David Lopez-Paz, “ mixup: Beyond empirical risk minimization”,in *International Conference on Learning Representations*, 2018.

[8] 李方, “基於深度視覺語義嵌入之零樣本學習” , 2019.

[9] Y.Xian, B.Schiele, Z.Akata,“Zero-shot learning – the good,the bad and the ugly”,in *CVPR*,2017.

[10] Yanwei Fu and Leonid Sigl,“Semi-supervised Vocabulary-informed Learning”,in *CVPR*,2016.

[11] Jie Song , Chengchao Shen , Yezhou Yang , Yang Liu , and Mingli Song,“Transductive Unbiased Embedding for Zero-Shot Learning”,in *CVPR*,2018.

[12] W.L.Chao, S.Changpinyo, B.Gong, F.Sha,“An Empirical Study and Analysis of Generalized Zero-Shot Learning for Object Recognition in the Wild”,in *ECCV*,2016.

[13] Long Chen, Hanwang Zhang, Jun Xiao, Wei Liu, Shih-Fu Chang,“Zero-shot visual recognition using semantics-preserving adversarial embedding networks”,in *CVPR*,2018.

[14] Yashas Annadani and Soma Biswas,“Preserving semantic relations for zero-shot learning”,in *CVPR*,2018.

[15] Huajie Jiang, Ruiping Wang, Shiguang Shan, and Xilin Chen,“Learning class prototypes via structure alignment for zero-shot recognition”,in *ECCV*,2018.

[16] Jin Li, Xuguang Lan, Yang Liu, Le Wang, and Nanning Zheng,“Compressing unknown images with product quantizer for efficient zero-shot classification”,in

CVPR,2019.

[17] M. Norouzi , T. Mikolov, S. Bengio, Y. Singer, J. Shlens, A. Frome, G. S. Corrado, J. Dean,“Zero-Shot Learning by Convex Combination of Semantic Embeddings”,in *ICLR*,2014.

[18] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, Marc’Aurelio Ranzato, T. Mikolov,“DeViSE: A Deep Visual-Semantic Embedding Model”,in *NIPS*,2013.

[19] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales,“Learning to compare: Relation network for few-shot learning”,in *CVPR*,2018.

[20] Yang Liu, Jishun Guo, Deng Cai, and Xiaofei He,“ Attribute attention for semantic disambiguation in zero-shot learning”,in *ICCV*,2019.

[21] Maunil R. Vyas, Hemanth Venkateswara, and Sethuraman Panchanathan,“Leveraging seen and unseen semantic relationships for generative zero-shot learning”,in *ECCV*,2020.

[22] Fang Li, Mei-Chen Yeh,“Generalized zero-shot recognition through image-guided semantic classification”,in *ICIP*,2021.

[23] P.Welinder, S.Branson, T.Mita, C.Wah, F.Schroff, S.Belongie, and P.Perona,“Caltech-UCSD Birds 200”,in *Caltech, Tech. Rep. CNS-TR-2010-001*,2010.

[24] G.Patterson and J.Hays,“Sun attribute database: Discovering, annotating, and recognizing scene attributes”,in *CVPR*,2012.

[25] A.Farhadi, I.Endres, D.Hoiem, and D.A.Forsyth,“Describing Objects by their Attributes”,in *CVPR*,2009.