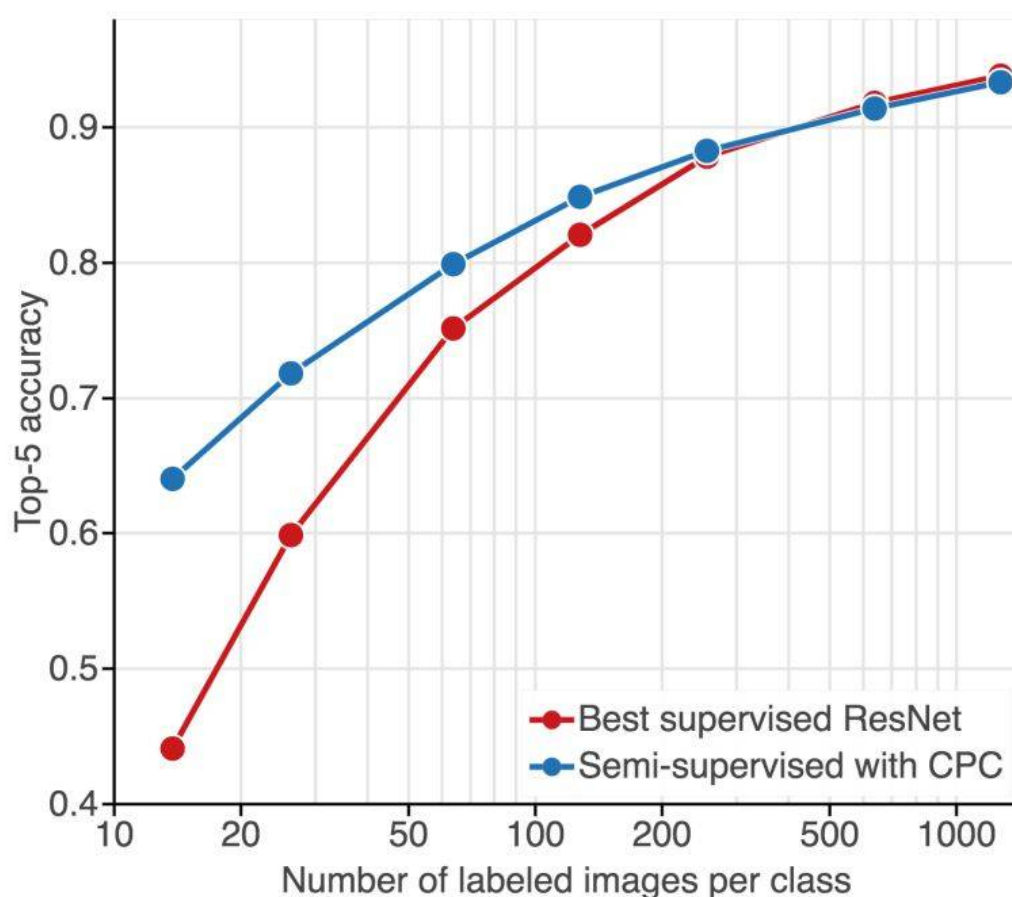


每類13張標註圖就可從頭學分類器，DeepMind新半監督模型超越AlexNet

© 2019-05-31

DeepMind 近期的一項研究利用對比預測編碼（Contrastive Predictive Coding，CPC）來解決這一難題，該方法是一種從靜止圖像中抽取穩定結構的無監督方法。得到的結果是一種表徵，使用該表徵和簡單的線性分類器在 ImageNet 上可實現優於其他方法的分類準確率，性能超越全監督 AlexNet 模型。即使給出少量標註圖像（每個類別 13 張標註圖像），該表徵也能保持強大的分類性能，Top-5 準確率超出當前最優的半監督方法 10%，超出當前最優監督方法 20%。



此外，研究發現，該無監督表徵可作為有效組件用於在 PASCAL-VOC 2007 數據集上的圖像檢測任務，且性能堪比使用全標註 ImageNet 數據集訓練得到的表徵。研究人員表示，希望該研究可以幫助在缺乏標註數據的現實視覺任務中，使用無監督表徵來替代監督表徵。

主要貢獻

具體而言，DeepMind 的這項研究改進了 CPC 模型 [49]，並基於此提出一種相對直接的方法。首先，研究人員對 CPC 進行架構優化，使其特徵編碼器擴展為較大的網絡，便於從無標註數據中提取更有用的信息，從而得到可以提高圖像分類準確率的特徵。具體來看，他們基於這些表徵訓練的線

性分類器在 ImageNet 上的 Top-1 準確率和 Top-5 準確率均優於全監督 AlexNet [37]。

其次，研究人員嘗試使用該表徵進行基於少量標籤（標註數據僅為 ImageNet 數據集的 1%）的分類。在該設置下，DeepMind 提出的半監督方法的 Top-5 準確率比當前最優監督方法高出 20%，比當前最優半監督方法高出 10%。與之前的半監督結果不同，該方法在標註樣本增加的情況下仍能保持強大性能，在使用完整 ImageNet 訓練集時，其性能堪比全監督性能，這表明該方法學得的特徵可以遷移到在線學習環境中。

第三，DeepMind 研究人員探索了該表徵的遷移學習能力。使用該無監督表徵作為特徵提取器在 PASCAL 2007 數據集上執行圖像檢測任務，得到的性能超越其他自監督遷移方法。重點是，這一性能結果接近專為監督遷移學習構建的方法。

最後，研究人員探索了不同的半監督學習方法，發現標準端到端精調方法並不一定是最優的。該研究發現，CPC 特徵可在未經重訓練的情況下使用，基於固定特徵訓練深度網絡。該方法的性能堪比精調後的模型性能，且其計算成本大大減少。這個結果很有意思，因為它復刻了自然語言處理領域的結果，NLP 領域中無監督特徵（如 word2vec [42] 和 BERT [13]）可在未經重新訓練的情況下在多個任務中取得強大性能，從而簡化訓練流程、減少計算成本。

什麼是對比預測編碼

對比預測編碼（CPC）是一種自監督方法，它基於之前觀測結果的表徵預測未來觀測結果的表徵，從而從序列數據中學習。當應用到圖像數據時，CPC 基於特定級別以上的圖像塊表徵預測該級別以下的圖像塊表徵（如圖 2 所示）。然後使用對比損失（contrastive loss）評估這些預測，在這一過程中網絡必須從一組無關的「負」表徵中正確分類出「未來」表徵。這避免了無效解，如用常數向量表示所有圖像塊（在使用均方誤差損失時會出現這種情況）。

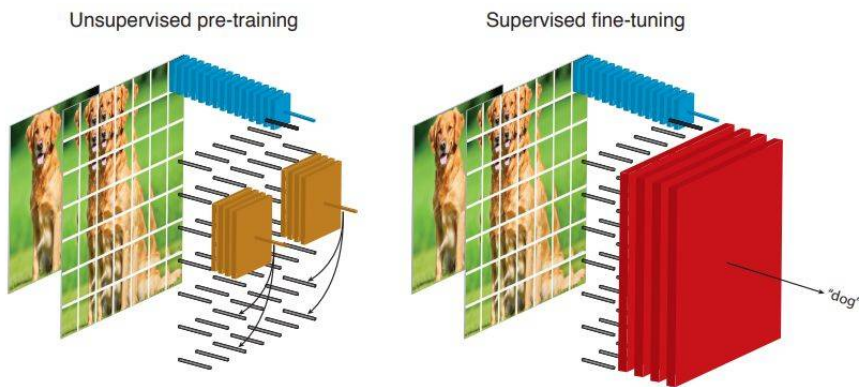


Figure 2. Overview of the framework for semi-supervised learning with Contrastive Predictive Coding. Left: unsupervised pre-training with a spatial prediction task. First, an image is divided into a grid of overlapping patches. Each patch is encoded independently from the rest with a feature extractor (blue) which terminates with a mean-pooling operation, yielding a single feature vector for that patch. Doing so for all patches yields a field of such feature vectors (wireframe vectors). Feature vectors above a certain level (in this case, the center of the image) are then aggregated with a context network (brown), yielding a row of context vectors which are used to linearly predict (unseen) feature vectors below. Right: using the CPC representation for a classification task. Having trained the encoder network, the context network is discarded and replaced by a classifier network (red) which can be trained in a supervised manner. For some experiments, we also fine-tune the encoder network (blue) for the classification task.

圖 2：使用 CPC 進行半監督學習的框架圖。左：使用無監督預訓練執行空間預測任務。右：使用 CPC 表徵執行分類任務。

方法詳解

利用 CPC 進行無監督學習

最近的研究表明，增加網絡容量和訓練規模可以提高性能。DeepMind 研究人員對現有 CPC 算法的第一個改進就是將網絡擴大。CPC 原來使用 ResNet-101-style 架構來表示每個圖像塊，他們則為該任務開發了一個更深、更廣的 ResNet。

然而，大型架構更加難以高效訓練。由於 CPC 必須在圖像塊上進行，這一問題變得更加嚴重。早期關於圖像塊語境預測的工作利用批歸一化來提高訓練速度。但 DeepMind 研究人員發現，對於大型架構來說，利用批歸一化得到的性能比較差。於是，他們利用層歸一化來彌補批歸一化的訓練效率。

有了更加高效、高容量的架構之後，研究者開發了一個有挑戰性的任務來對其進行訓練。他們首先通過向上預測（即對空間位置較低的圖像塊進行聚合以預測空間位置較高圖像塊的表徵）和向下預測（CPC 最初只使用向下預測）來加倍每張圖像中的監督信號。這兩種方向的預測使用不同的語境網絡。研究人員發現，額外的圖像塊增強可以帶來顯著的性能提升。首先，他們利用了 [14] 中的「color dropping」方法，該方法在每個圖像塊中隨機丟棄三個顏色通道的兩個。他們隨機對圖像塊進行水平翻轉。他們還隨機將一個完整的圖像塊裁剪為 56x56 大小的多個圖像塊，並將它們重新拼為原來的大小，從而在空間上對其進行抖動。

增加任務的複雜度之後，CPC 的目標變得非常困難，即使對於如此高容量的模型來說也是如此。在實踐中，研究者發現增加任務難度竟然提高（而不是降低）了模型在下游任務中的性能。實際上，如果網絡可以學習利用低級圖案（如變化緩慢的顏色，或圖像塊之間的連續直線）來完成任務，那麼該網絡就無需學習任何語義上有意義的語境。通過增加圖像塊之間的低級可變性，研究者移除了這些低級特徵，同時也加大了任務難度，迫使網絡通過抽取高級特徵來解決問題。

利用 CPC 進行半監督學習

研究者探索了兩種方式來將 CPC 與有監督分類任務結合在一起。第一種固定範式包括專為 CPC 優化一個特徵提取器 f_θ 。然後固定其參數並優化分類器 g_ϕ ，以區分上述特徵提取器的輸出。從公式上來說，給定一個含有 N 張圖像的數據集 $\{x_n\}$ ，得到：

$$\theta^* = \arg \min_{\theta} \frac{1}{N} \sum_{n=1}^N \mathcal{L}_{\text{CPC}}[f_{\theta}(x_n)].$$

給定一個含有 M 張標註圖像的數據集 $\{x_m, y_m\}$ （這個數據集可能要小得多），得到：

$$\phi^* = \arg \min_{\phi} \frac{1}{M} \sum_{m=1}^M \mathcal{L}_{\text{Sup}}[g_{\phi} \circ f_{\theta^*}(x_m), y_m].$$

研究者還探索了一種微調機制，這種機制允許特徵提取器適應有監督目標。確切點說，研究者利用在上述學習階段得出的解 θ^* 和 ϕ^* 初始化特徵提取器和分類器，併為有監督目標微調整個網絡。為了確保特徵提取器不偏離 CPC 指定的解太多，研究者應用了更小的學習率和早停。

但 CPC 要求特徵提取器 f_θ 獨立應用在疊加的圖像塊上，在半監督學習階段，它可以直接應用在整個圖像上。這將整體計算量減少了 2/3-3/4，因此可以加速訓練並減少內存佔用。為了減小圖像塊上的無監督學習與整幅圖像上的監督微調之間的域不匹配，研究者在所有的卷積中使用了對稱填

充，並在無監督預訓練過程中使用了空間抖動。

針對圖像分類任務對模型進行訓練時，分類器 g^{ϕ} 是一個 11-block ResNet 架構，擁有 4096 維特徵圖和 1024 維瓶頸層。監督損失函數 L_{Sup} 是模型預測和圖像標籤之間的交叉熵。當對模型進行圖像檢測訓練時，研究者使用了 Faster-RCNN 架構和損失函數，沒有做任何修改。

實驗

在 ImageNet 上的分類性能

該研究首先探索了該模型線性分割圖像類別的能力，這是無監督表徵學習的標準基準測試。實驗結果表明，改進版的 CPC 架構性能顯著超過之前公開方法的性能結果，如下表 1 所示：

Method	Top-1	Top-5
Motion Segmentation (MS) [50]	27.6	48.3
Exemplar (Ex) [17]	31.5	53.1
Relative Position (RP) [14]	36.2	59.2
Colorization (Col) [69]	39.6	62.5
Combination of MS + Ex + RP + Col [15]	-	69.3
CPC [49]	48.7	73.6
Rotation + RevNet [36]	55.4	-
CPC (ours)	61.0	83.0

表 1：該研究提出的方法與其他自監督方法的線性分割能力對比。在所有實驗中，特徵提取器都以無監督的方式進行優化，線性分類器使用 ImageNet 數據集中的所有標籤進行訓練。

接着，研究人員在 ImageNet 數據集上評估全監督網絡在具備不同數量標註訓練數據時的性能。

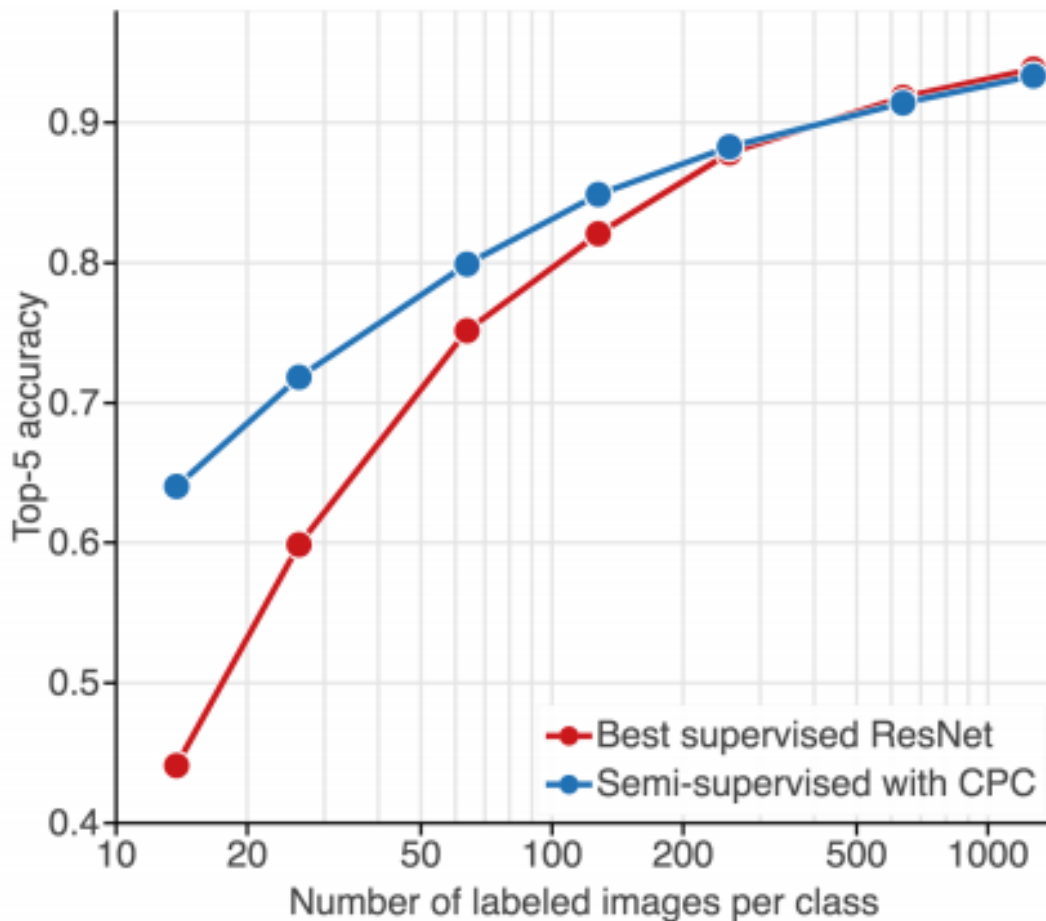


圖 1：分類準確率 vs. 標註樣本數量。監督方法（紅色線）的性能隨着標註數據數量下降而急劇下滑；使用大量無標註樣本（藍色線）對這些方法進行正則化可以極大地緩解性能的下降。

如圖 1 所示，隨着數據量下降，全監督模型過擬合愈發嚴重。儘管研究者相應地提升了正則化，但該模型的性能從 93.83% 的準確率下降到了 44.10%（前者是在整個數據集上進行訓練的結果，後者是在 1% 的數據上進行訓練的結果，見圖 1 和圖 3 的紅線）。

最後，研究者對比了他們提出的方法和全監督基線模型。研究者在整個無標註 ImageNet 數據集上預訓練特徵提取器，學習分類器，並使用標註圖像的子集進行模型精調。圖 1 和圖 3 中的藍線展示了該方法的結果。

之後，研究者對比了他們提出的方法和其他半監督學習方法。下圖 3 展示了這些方法沒能顯著改善基線監督模型，CPC 方法是唯一超過監督學習模型的方法。

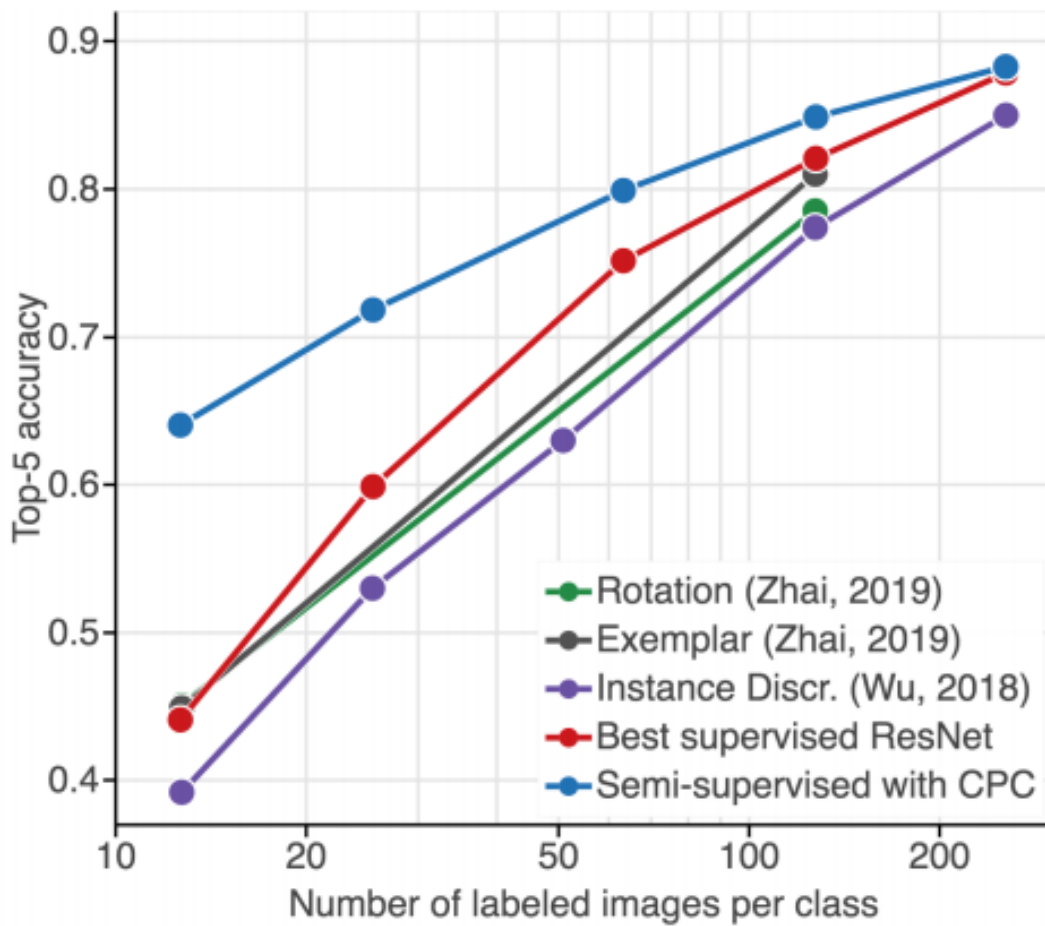


圖 3：該研究提出方法與使用自監督學習和監督精調模型的半監督學習方法的對比。藍色線：使用 CPC 的半監督學習。紫色線：使用 *instance discrimination* 的半監督學習 [64]。綠色線：使用 *rotation prediction* 的半監督學習 [68]。灰色線：使用 *exemplar learning* 的半監督學習。紅色線：監督基線方法。

Labeled data Method	1% Top-5 accuracy	10%
Supervised baseline	44.10	82.08
<i>Methods using label-propagation:</i>		
Pseudolabeling [68]	51.56	82.41
VAT [68]	44.05	82.78
VAT + Entropy Minimization [68]	46.96	83.39
Unsup. Data Augmentation [65]	-	88.52
Rotation + VAT + Ent. Min. [68]	-	91.23
<i>Methods only using representation learning:</i>		
Instance Discrimination [64]	39.20	77.40
Exemplar [68]	44.90	81.01
Exemplar (joint training) [68]	47.02	83.72
Rotation [68]	45.11	78.53
Rotation (joint training) [68]	53.37	83.82
CPC (ours)	64.03	84.88

表 2：在使用 1% 或 10% 標註數據時的半監督學習方法性能對比。表徵學習方法以無監督方式學習表徵，並將表徵用於分類。分類器僅考慮標註樣本，且僅侷限於監督式方法。

遷移到 PASCAL 後的圖像檢測性能

有用的無監督學習方法所訓練的表徵可以很好地遷移到新數據集和任務。爲了調查 CPC 所學表徵是否具備該能力，研究人員在 PASCAL 數據集上評估其圖像檢測性能。下表 3 展示了該方法與其他方法的對比結果：

Method	mAP
<i>Transfer from labeled ImageNet:</i>	
Supervised - ResNet-152	74.7
<i>Transfer from unlabeled ImageNet:</i>	
Exemplar (Ex) [17]	60.9
Motion Segmentation (MS) [50]	61.1
Colorization (Col) [69]	65.5
Relative Position (RP) [14]	66.8
Combination of	
Ex + MS + Col + RP [15]	70.5
Deep Cluster [8]	65.9
Deeper Cluster [9]	67.8
CPC - ResNet-101	70.6
CPC - ResNet-170	72.1

表 3：在 PASCAL 2007 數據集上執行圖像檢測任務時，不同遷移方法的性能對比結果。第一類模型從無標註 ImageNet 數據中學習，並針對 PASCAL 圖像檢測任務進行精調。第二類模型在遷移之前先基於整個標註 ImageNet 數據集學習。所有結果均以 mAP 來衡量。

參考鏈接：<https://arxiv.org/pdf/1905.09272.pdf>

文章來源：機器之心 (<https://www.jiqizhixin.com/articles/2019-05-31-9>)

 喜歡這篇文章嗎？快分享吧！