

Class-incremental Learning via Deep Model Consolidation (WACV 2020)

ChangJun Ahn
Hanbat National University Dept. of Electronic Engineering
qkqk1009@gmail.com
Aug 1, 2023

논문소개: Class-incremental Learning via Deep Model Consolidation (WACV 2020)

- 새로 추가된 클래스 집합에 대한 학습 목표를 적용할 때의 원래 클래스 집합에 대한 성능이 급격하게 저하되는 “**Catastrophic Forgetting**” 현상을 초래합니다.
- Incremental Learning 은 이전 클래스나 새로운 클래스에 대해 **biased** 된 model을 생성하는 경향이 있습니다.
- 이를 해결하기 위해서 본 논문에서는 “**Deep Model Consolidation(DMC)**” 이라는 **Class incremental learning paradigm** 제안합니다.

Novel ideas

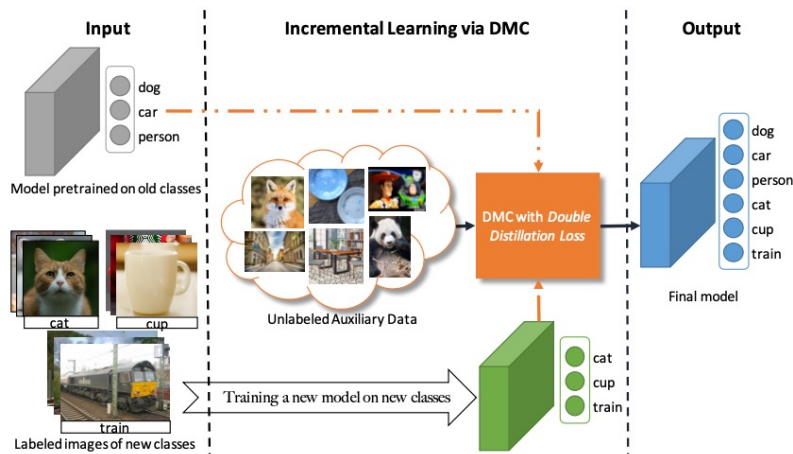
1. 먼저 새로운 클래스에 대해서만 따로 훈련을 시킵니다.
2. 두개의 모델을 결합, 이 모델들은 두개의 서로 다른 클래스 집합의 데이터를 기반으로 훈련합니다.
3. 두 기존의 모델은 publicly available unlabeled auxiliary data를 활용하여 통합됩니다.



“**Knowledge Distillation** 을
일반화한 **Double Distillation** 제시”

Proposed

- 분류된 데이터를 이용하여 새로운 클래스에 대해 따로 모델을 훈련합니다.
- **old class** 와 **new class**의 모델을 **publicly available unlabeled auxiliary data**를 활용하여 **double distillation**으로 훈련합니다



- DMC는 legacy data의 접근의 어려움을 unlabeled auxiliary data로 극복하며, 이전에 전달 가능한 풍부한 표현들을 활용하여 Incremental learning을 촉진할 수 있습니다.
- DMC는 intrinsic bias에 의한 **information asymmetric, over regularization**을 훈련에서 제거하고, 제안된 **double distillation**으로 두개의 teacher model로 부터 최종 student model을 **동시에** 학습한다.
- 새로운 클래스의 훈련 데이터 대신 **auxiliary data**를 사용함으로써 student model이 두 teacher model로 부터 전달된 지식을 **편향 없이** 흡수할 수 있도록 보장합니다.

DMC for image classification

1) Single compact model

- 새로운 클래스에 대해 사용 가능한 훈련 데이터 D_{new} 를 사용하여 standard softmax cross entropy loss로 CNN 모델 f_{new} 학습합니다.

2) Training objective for consolidation

- \mathcal{U} 는 unlabeled auxiliary train data를 나타냅니다.

3) Define double distillation loss

- double distillation loss는 L_{dd} 로 정의 됩니다.
- 클래스 차원에 대해 각 모델에 평균을 빼 정규화를 합니다.

4) Bias calibration

- 두 개의 클래스에 대한 **bias 보정** 단계로 작용하며, 이 방법은 두 모델이 생성하는 logit의 scale을 통일하고, 클래스 간의 크기를 유지하여 **symmetric**한 information flow를 유지합니다.

defined as :

$$f(x; \Theta)[j] = \begin{cases} f_{old}(x; \Theta_{old})[j], & 1 \leq j \leq s \\ f_{new}(x; \Theta_{new})[j], & s < j \leq t \end{cases}, \forall x \in \mathcal{I} \quad (1)$$

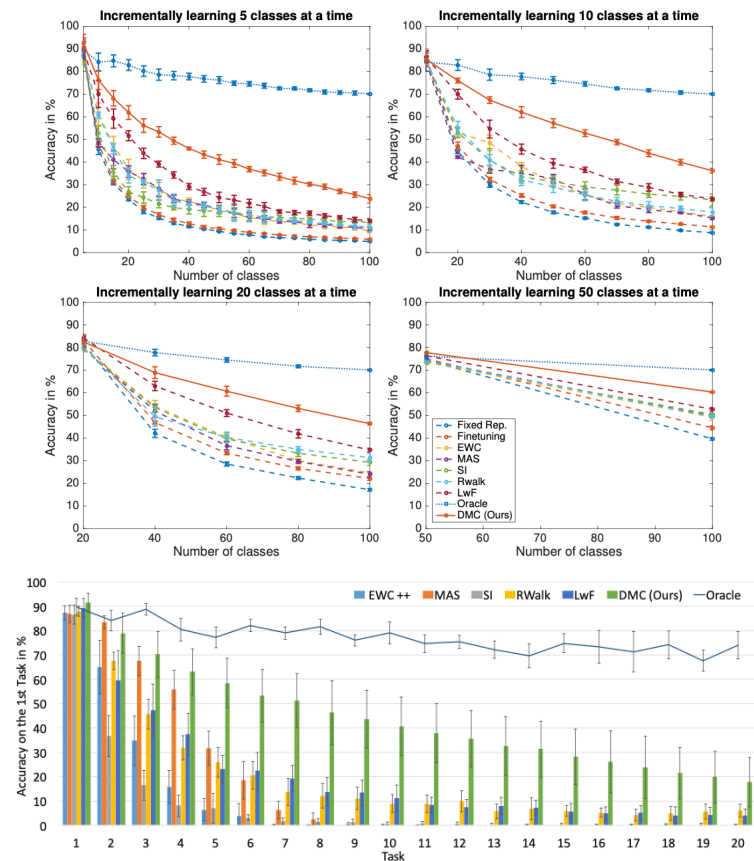
$$\min_{\Theta} \frac{1}{|\mathcal{U}|} \sum_{x_i \in \mathcal{U}} L_{dd}(\mathbf{y}_i, \hat{\mathbf{y}}_i), \quad (2)$$

$$L_{dd}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{t} \sum_{j=1}^t (y^j - \hat{y}^j)^2, \quad (3)$$

$$\hat{y}^j = \begin{cases} \hat{y}^j - \frac{1}{s} \sum_{k=1}^s \hat{y}^k, & 1 \leq j \leq s \\ \hat{y}^j - \frac{1}{t-s} \sum_{k=s+1}^t \hat{y}^k, & s < j \leq t \end{cases} \quad (4)$$

Experiments

- Datasets 는 CIFAR-100으로 진행하고, 100개의 class를 한번 학습 할 때 5, 10, 20, 50 으로 나누어 진행합니다.
- 모든 실험에는 ResNet-32로 진행되었습니다.
- 다음과 같은 방법을 사용하며 최첨단 성능이 무엇인지 비교했습니다.
(DMC, Oracle, Fixed Rep, Finetuning, EWC++, LwF, SI, MAS, RWalk)
- DMC는 클래스 수가 증가하더라도 안정적인 성능을 보이며, 클래스 값이 변하더라도 DMC의 성능은 크게 영향을 받지 않습니다.
- 다른 regularization들은 기존 클래스와 새로운 클래스 간의 **information asymmetric**으로 인해 incremental learning이 증가함에 따라 성능이 안정적이지 못합니다.



Conclusion

- 본 논문에서는 새로운 class incremental learning paradigm인 **DMC**를 제시합니다.
- Knowledge Distillation의 기본 아이디어를 일반화하고, class incremental learning 을 가능하게 하기 위해 **double distillation loss**를 제안합니다.
- 새로운 double distillation training 목표를 활용하여 DMC는 기존 데이터의 저장을 필요로 하지 않습니다.
- DMC는 준비된 **unlabeled auxiliary data**를 활용하여 두 개의 독립적으로 훈련된 모델을 **consolidation** 시킵니다.
- DMC를 통해 intrinsic bias에 의해 **information asymmetric, over-regularization**을 훈련에서 제거합니다.
- DMC는 다른 방법들과 달리 클래스 수가 증가하여도 안정적인 성능을 보이며, 다른 regularization 들은 클래스 간 **information asymmetric** 으로 인해 성능이 저하 하게됩니다.