Journal of Korea Multimedia Society Vol. 21, No. 2, February 2018(pp. 98-107)
https://doi.org/10.9717/kmms.2018.21.2.098

골 성숙도 판별을 위한 심층 메타 학습 기반의 분류 문제 학습 방법

민정원†, 강동중††

Deep Meta Learning Based Classification Problem Learning Method for Skeletal Maturity Indication

Jeong Won Min†, Dong Joong Kang††

ABSTRACT

In this paper, we propose a method to classify the skeletal maturity with a small amount of hand wrist X-ray image using deep learning-based meta-learning. General deep-learning techniques require large amounts of data, but in many cases, these data sets are not available for practical application. Lack of learning data is usually solved through transfer learning using pre-trained models with large data sets. However, transfer learning performance may be degraded due to over fitting for unknown new task with small data, which results in poor generalization capability. In addition, medical images require high cost resources such as a professional manpower and mcuh time to obtain labeled data. Therefore, in this paper, we use meta-learning that can classify using only a small amount of new data by pre-trained models trained with various learning tasks. First, we train the meta-model by using a separate data set composed of various learning tasks. The network learns to classify the bone maturity using the bone maturity data composed of the radiographs of the wrist. Then, we compare the results of the classification using the conventional learning algorithm with the results of the meta learning by the same number of learning data sets.

Key words: Meta Learning, Skeletal Maturity Indicator, Classification

1. 서 론

최근 영상 인식, 자연어 번역 등 다양한 분야에 딥러닝 기술을 접목하여 기술의 발전이 이루어지면서 여러 분야에서 딥러닝에 대한 관심이 높아지고 있다. 특히 영상 인식의 경우 공정의 결함검사, 자율주행 자동차 및 번호판 인식[1], 의료 영상에서의 병변 검출 등 많은 분야에서 사용되며 여러 분야에서 뛰어난 성능을 보이고 있다. 이러한 딥러닝 기반의 모델을 학습하기 위해서는 대량의 데이터 셋이 필요하다. 특히 지도학습(supervised learning)을 위해서는 각 데이터에 대한 분류 정보를 담는 라벨링(labeling) 과정을 필요로 한다. 딥러닝을 이용한 기술이 좋은 성능을 갖기 위해서는 일반적으로 수천, 수만 장의 학습데이터가 필요하며 데이터 셋의 구성에 따라 학습 성능이 달라지기 때문에 양질의 데이터

※ Corresponding Author : Dong Joong Kang, Address: (46241) 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, Republic of Korea, TEL : +82-51-510-2356, FAX : +82-51-514-1118, E-mail : djkang@pusan.ac.kr
Receipt date : Jan. 15, 2018, Approval date : Jan. 22, 2018
† Dept. of Control, Automation and Systems Eng., Graduate School, Pusan National University
(E-mail : jwmin@pusan.ac.kr)
†† Dept. of Control, Automation and Systems Eng., Graduate School, Pusan National University
※ This research was supported by Basic Research Support Program (2years) through the Pusan National University (PNU).

셋을 확보해야 한다. 또한 이러한 대량의 데이터 처리를 위해 별도의 병렬 연산장치가 필요하고 연산에도 오랜 시간이 걸린다. 따라서 딥러닝 기술을 적용하고자 하는 많은 문제에는 대량의 학습데이터 수집과 모델 학습에 어려움이 존재한다.

인공신경망 모델을 학습하는 과정에서 학습데이터가 부족한 경우 일반적으로 전이 학습(Transfer Learning) 방법을 사용한다. 전이 학습은 다양한 종류의 데이터를 포함하는 대량의 데이터 셋을 이용하여 사전에 모델을 학습한 후 실제 학습하고자 하는 소량의 데이터 셋을 이용해 신경망의 가중치 중 일부만을 새로 학습하는 방법이다. 영상 인식에서 많이 사용되는 컨볼루션 신경망(Convolution Neural Network; CNN)의 경우 특징 추출을 담당하는 컨볼루션 층(Convolution Layer)의 가중치는 학습하지 않고 회귀분석을 담당하는 완전 연결 층(Fully Connected Layer)의 가중치만을 새로운 데이터를 이용해 학습한다. 하지만 전이 학습에도 적지 않은 양의 학습 데이터가 필요하며 데이터 량이 적을 경우 과적합(Overfitting) 현상이 발생하여 모델의 일반화 성능이 떨어진다. 따라서 기존의 경험을 활용하여 적은 데이터량으로 새로운 문제에 쉽게 적응할 수 있는 학습 방법에 대한 연구가 이루어지고 있다.

이러한 분야 중 한 가지가 메타 학습(Meta Learning)이다. 메타 학습이란 기존의 학습 경험을 이용하여 어떤 모델이 새로운 학습 문제에 유연하고 빠르게 적응할 수 있도록 학습시키는 방법이다. 메타 학습은 주로 메타 학습자(Meta learner)를 통해 모델이 학습을 잘 수행할 수 있는 방법 자체를 학습하는 것을 목표로 한다. Ravi[2]는 가중치를 갱신하는 최적화 함수를 학습하여 새로운 과제에 알맞은 함수를 찾도록 했으며 Finn[3], Li[4]는 다양한 과제로 이루어진 메타 데이터 셋을 이용하여 모델의 초기 가중치를 학습 오차에 민감하게 반응하도록 학습한 후, 학습되지 않은 새로운 과제에 대해서도 적은 양의 데이터만으로 좋은 일반화 성능을 내도록 했다.

본 논문에서는 이러한 메타 학습 기법을 데이터 확보가 어려운 골 성숙도 판별(Skeletal Maturity Indicator : SMI)에 적용한다. 골 성숙도는 사람의 생물학적 나이를 판단하는 지표이며 의학적 진단의 중요한 기준으로 사용된다. 일반적으로 수완부의 X-ray 영상을 이용해 골 성숙도를 판단하나 이를 위해서는 전문가 인력이 필요할 뿐만 아니라 환자에 따라 많은 예외가 존재함에 의해 측정 기준이 모호하므로 정확한 분류가 어렵다. 때문에 골 성숙도 자동 판단 시스템을 위한 많은 연구가 진행되고 있으나 이는 여전히 어려운 문제이다.

따라서 본 논문에서는 사전에 미리 메타 모델을 학습하여 소량의 수완부 X-ray 영상으로 골 성숙도를 판별하는 네트워크를 제안하고자 한다. 또한, 개인 정보 보호 등으로 인해 대량의 데이터 확보가 어려운 의료 분야의 메타 학습을 이용한 데이터 분류의 가능성을 제시한다. 제안하는 시스템은 부류 별 20장의 수완부 X-ray 영상을 학습하여 5개의 부류를 분류하는 것을 목표로 한다. 본 논문에서 사용하는 메타 학습 방법인 MAML(Model-Agnostic Meta-Learning)[3]은 다양한 학습과제로 이루어진 메타 데이터 셋을 이용하여 모델을 사전에 학습한 후 이 모델을 이용하여 학습되지 않은 새로운 데이터 셋에 대해 적은 양의 데이터량 만으로 비교적 좋은 성능을 보인다. 학습에 사용되는 신경망의 구조는 영상 판별에 뛰어난 성능을 보이는 컨볼루션 신경망(Convolution Neural Networks)[5]를 사용한다. 컨볼루션 신경망은 특징 추출과 분류를 동시에 수행하여 사용자가 특징을 모델링 하는 과정이 필요하지 않으며 영상 인식, 검출에서 아주 뛰어난 성능을 보이는 방법이다. 본 논문에서는 공공 데이터 셋으로 제공되는 TinyImageNet[6], Cifar100[7]데이터 셋을 사용하여 사전에 메타 모델을 학습한 후 이 모델을 사용해 각 단계별 20장의 데이터로 5개의 성숙 단계를 분류한다. 이때 골 성숙도를 판별하는데 중요한 영역인 손가락 마디 6개를 ROI(Region of Interest)로 잘라내어 학습 데이터로 사용한다. 비교실험으로는 학습에 사용할 데이터 량이 작을 때 일반적으로 사용하는 전이학습(Transfer Learning) 방법과 임의로 초기화된 모델을 처음부터 학습하는 scratch 학습 방법으로 같은 양의 데이터를 학습하여 결과를 비교한다. 두 가지 방법에 대해 메타 학습을 적용한 결과와 일반적으로 학습한 결과를 각각 실험하고 비교한다.

본 논문은 다음과 같이 구성된다. 2장에서 학습 모델의 기반이 되는 컨볼루션 신경망에 대한 이론과 메타 학습을 적용하고자 하는 골 성숙도 판별 문제에 대해 설명한다. 3장에서는 본 논문에서 사용하는 메타 학습 기법 중 하나인 Model-Agnotsic Meta-

Learning(MAML) 알고리즘을 소개하고, 4장에서는 골 성숙도 판별 문제에 적용하는 방법과 데이터 준비 과정에 대해 설명한다. 이때 학습 데이터가 적을 경우 사용하던 기존의 방법인 전이 학습과 메타 학습의 결과를 비교한다. 마지막으로 5장에서 본 연구의 결론에 대해 서술한다.

2. 심층 메타 학습

본 논문에서 사용하는 메타 학습 방법인 MAML (Model-Agnostic Meta-Learning)은 메타 학습자(Meta Learner)가 여러 개의 학습자(Learner)를 이용해 다양한 데이터에 대한 학습 경험을 쌓은 후 새로운 데이터에 최적화된 초기 가중치를 찾는 방법이다. 여기서 메타 학습자와 학습자는 주어진 분류 문제(Task)를 해결하기 위한 가중치의 집합을 나타낸다. 학습자는 하나의 학습 문제에 최적화된 가중치를 학습하고 메타 학습자는 여러 학습자로부터 다양한 학습 문제에 최적화된 가중치에 대한 학습 결과를 전달받는다.

Fig. 1은 이러한 MAML 방법과 일반적인 학습 방법인 Base Learning의 학습 과정을 나타낸 다이어그램이다. θ는 가중치, 굵은 화살표는 θ의 업데이트 방향을 나타낸다. Base Learning의 $\triangle L$은 손실 함수를 통해 계산되는 손실의 그래디언트(gradient)를 나타내며 θ^*는 손실의 그래디언트를 반영하여 갱신될 가중치를 나타낸다. 이때 기존의 학습 방법인 Base Learning은 하나의 분류 문제를 학습하기 때문에 그 문제에만 최적화된 가중치를 갖게 된다. 반면 MAML방법은 여러 개의 분류 문제를 여러 개의 학습자를 이용해 동시에 학습한다. 세 개의 분류 문제를 동시에 학습했을 경우 각각의 문제는 손실의 그래디언트 $\triangle L_1, \triangle L_2, \triangle L_3$을 고려한 가중치 업데이트 방향 $\theta_1^*, \theta_2^*, \theta_3^*$를 가지게 되며 이는 Fig. 1에서 점선으로 나타내었다. MAML 방법에서 메타 학습자 θ^M은 여러 개의 학습자에 대한 손실들을 종합적으로 고려한 그래디언트 업데이트 방향을 찾도록 학습된다. 따라서 이후 학습한 경험이 없는 새로운 분류 문제를 접했을 때 이 문제에 최적화된 가중치를 출력할 수 있다.

Fig. 2는 메타 학습자와 각 과제를 학습하는 학습자들의 그래디언트 업데이트(gradient update) 과정

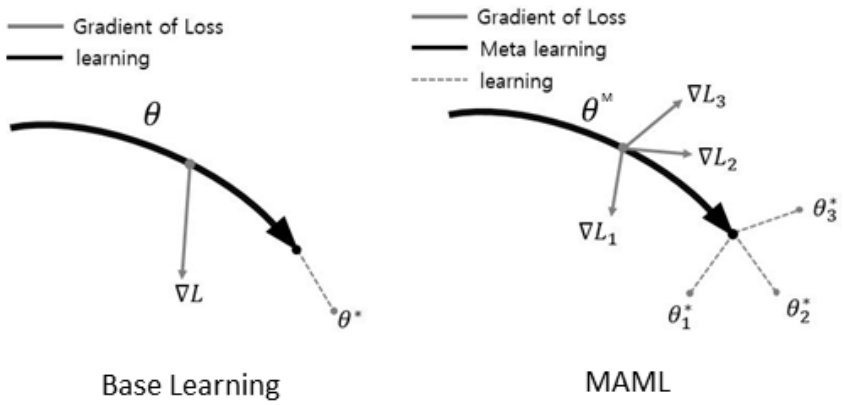

Fig. 1. Learning process for general Base Learning and MAML.

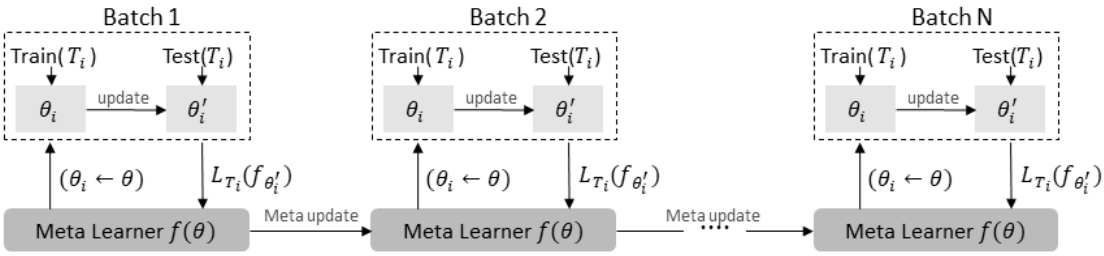

Fig. 2. Flowchart of learning process of Model-Agnostic Meta Learning.

을 나타낸 흐름도이다. 메타 학습자를 학습하는 데 사용되는 메타 데이터 셋은 학습 샘플과 테스트 샘플로 나뉜다. Fig. 2의 T_i는 임의의 분류 문제를 의미하며 이 분류 문제는 학습 샘플(Train(T_i)), 검증 샘플(Test(T_i))로 나뉜다. 메타 학습에서는 한 분류 문제를 한 개의 데이터로 취급하기 때문에 학습 샘플과 검증 샘플은 분류 문제에 포함되는 여러 장의 영상으로 구성되어 있다. 예를 들어 분류하고자 하는 부류가 2개이고, 각 부류 당 학습 샘플이 5장, 검증 샘플이 15장일 경우 Task i의 학습 샘플은 $(5+15)\times 2=40$장의 영상으로 구성된다.

메타 학습자 $f(\theta)$는 전달 받은 학습 샘플을 이용해 1번의 그래디언트 업데이트를 수행하여 새로운 가중치 $\theta_i^{'}$를 생성한다. 이 가중치는 T_i의 학습 샘플을 한 번 학습한 결과이며 학습자 $f(\theta'_i)$에게 전달되는 T_i에 대한 초기 가중치이다. 학습자는 이 가중치를 전달받는 동시에 메타 학습자의 그래디언트 업데이트에 사용되었던 T_i의 학습 샘플을 다시 전달받는다. 학습자는 이 학습 샘플을 이용해 여러 번의 그래디언트 업데이트를 수행하여 입력된 학습 샘플을 학습한다. 메타 학습자와 학습자의 그래디언트 업데이트는 식 (1)과 같으며 Fig. 2에서 'update' 과정에 해당한다.

$$\theta_i^{'} = \theta - \alpha \nabla_\theta L_{T_i}(f_\theta) \tag{1}$$

이렇게 학습된 학습자는 분류 문제 T_i 학습에 최적화된 가중치를 갖는다. 이후 T_i의 검증 샘플을 전달받은 학습자는 학습된 가중치를 이용해 테스트 샘플의 부류를 예측하여 실제 부류에 대한 손실 $L_{T_i}(f_\theta)$을 계산한다. $\nabla_\theta L_{T_i}(f_\theta)$는 모델이 예측한 부류와 실제 부류의 차이에 대한 손실의 그래디언트를 의미하며 손실함수 L은 Cross Entropy 함수를 이용한다. 이는 식 (2)에 나타내었다. x^j, y^j는 입력, 출력의 라벨을 의미하며 T_i는 task, f는 모델을 의미한다.

$$L_{T_i}(f_\phi) = \sum_{x^j, y^j \sim T_i} y^j \log f_\phi(x^j) + (1-y^j)\log(1-f_\phi(x^j)) \tag{2}$$

이후 손실에 대한 그래디언트는 메타 학습자에게 전달된다. 메타 학습자는 이렇게 자신이 학습자에게 제공해주었던 초기 가중치의 성능에 대한 손실의 그래디언트를 전달받는다. 한 배치에 소속된 여러 Task에 대해 손실을 되돌려 받는 과정을 반복한 메타 학습자는 모든 손실의 합을 이용해 자신의 가중치 θ를 업데이트한다. 이 과정은 식 (3)과 같으며 Fig. 2에서 'meta update' 과정에 해당한다.

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta_i^{'}}) \tag{3}$$

이러한 방법의 학습을 통해 분류 문제에 최적화된 가중치를 갖게 된 학습자는 문제의 검증 샘플에 의해 성능이 평가된다. 검증 샘플에 대해 계산된 손실에 대한 그래디언트 $\nabla_\theta L_{T_i}(f_{\theta_i^{'}})$는 메타 학습자에게 전달된다. 메타 학습자는 한 배치(Batch)에 포함되는 모든 분류 문제들에 대해 학습자들에게 초기 가중치를 제공하고 검증 샘플에 대한 손실을 전달받는다. 이때 배치란 여러 개의 데이터를 하나로 묶은 데이터의 집합을 의미한다. 신경망 학습 과정에서는 일반적으로 데이터를 배치 형태로 학습하여 일반화 성능을 높인다. 메타 학습에서는 하나의 분류 문제가 하나의 데이터이므로 여러 개의 과제가 묶인 형태의 배치를 학습한다. 배치에 대한 학습과 검증이 끝난 후 메타 학습자는 전달받은 손실을 이용해 가중치 θ를 update한다. 이때 가중치는 메타 학습자가 생성한 초기 가중치가 학습을 잘 수행하였는가에 대한 결과를 통해 학습되므로 결국 입력되는 데이터에 대해 최적화된 초기 가중치를 생성하는 방향으로 학습이 진행된다.

한편 식 (1)의 α는 과제 모델을 학습할 때 사용되는 학습율(learning rate)이다. 학습율이란 가중치가 gradient 방향에 대해 얼마나 이동할지를 나타내는 값이다. MAML 방법은 학습되는 데이터량이 작으므로 학습 오차에 민감하게 반응하기 위해 비교적 큰 학습율 값을 사용한다. 식 (3)의 β도 마찬가지로 학습율을 나타내며, 학습자에 대한 학습율인 α와 달리 메타 학습자의 가중치를 업데이트할 때 사용한다.

이렇게 적은 양, 적은 수의 업데이트를 통해 학습된 학습자 $f_{\theta_i^{'}}$를 검증하고, 그 결과는 메타 모델의 학습 오차로 사용된다. 결국 메타 모델은 하나의 데이터 - 분류 문제를 학습하고 검증한 결과를 메타 모델에 반영하여 학습을 더 잘 수행하는 방향으로 가중치를 갱신한다. 이러한 과정을 통해 다양한 부류로 이루어진 여러 과제에 대해 학습을 수행함으로써 메타 모델은 새로운 학습과제의 학습 오차에 민감하게 반응하는 초기 가중치를 갖게 된다.

3. 수완부의 메타학습

3.1 알고리즘 개요

본 논문에서는 사전에 미리 메타 모델을 학습하여 소량의 수완부 X-ray 영상으로 골 성숙도를 판별하는 네트워크를 제안하고자 한다. 제안하는 골 성숙도 판별 알고리즘의 흐름도를 Fig. 3에 나타내었다. 메타 학습을 골 성숙도 판별 문제에 적용하기 위해서는 먼저 학습된 메타 학습자가 필요하다. 메타 학습자는 2장의 방법으로 학습되며 학습이 완료된 메타 학습자는 실제 해결하고자 하는 문제에 적용될 수 있다.

먼저 메타 데이터 셋을 이용하여 메타 학습자를 학습하며, 이는 2절의 메타 학습 과정과 같다. 다양한 부류로 구성된 분류 문제를 전달 받은 메타 학습자는 한 번의 그래디언트 업데이트 과정을 통해 학습자에게 전달할 초기 가중치를 생성한다. 학습자는 전달받은 가중치를 초기값으로 하여 학습 샘플을 이용해 여러 번의 그래디언트 업데이트 과정을 반복한다. 그 결과 학습자의 가중치는 해당 분류 문제에 최적화된 가중치를 갖게 된다. 이후 해당 분류 문제의 검증 샘플을 이용하여 실제 부류와 학습자를 통해 예측된 값 사이의 손실과 그 손실에 대한 그래디언트를 계산한다. 이 그래디언트는 다시 메타 학습자에게 전달되며 메타 학습자는 이 그래디언트를 이용하여 자신의 가중치를 업데이트한다. 그 결과 메타 학습자는 처음에 입력받은 학습 샘플에 대해 빠르게 적응하고 좋은 성능을 내는 초기 가중치를 생성할 수 있는 방향으로 학습된다. Fig. 3의 과정 ①이 이에 해당한다.

Fig. 3의 과정 ②는 학습된 메타 학습자를 실제 해결하고자 하는 분류 문제에 적용하는 과정이다.

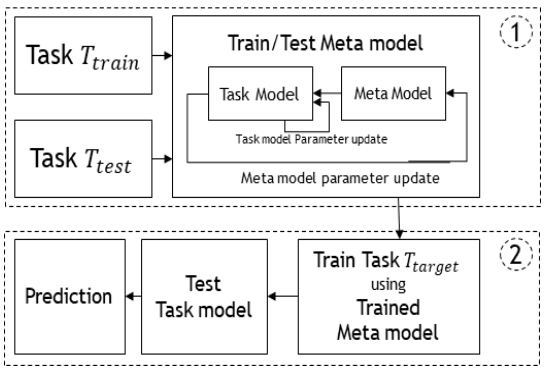

Fig. 3. Overall system flow chart.

Task T_{target}은 실제 해결하고자 하는 분류 문제에 해당하는 영상 데이터이다. 메타 학습자는 학습자에게 초기 가중치를 제공하기 때문에 신경망의 구조가 학습자와 동일해야하므로 Task T_{target}역시 2개의 부류로 구성되어야 한다. Task T_{target}의 학습 샘플은 메타 학습자를 학습하기 위해 사용했던 메타 데이터 셋에는 포함되지 않은 영상들이며 본 연구에서는 수완부 영상 데이터 셋에 해당한다. 따라서 메타 학습자는 이 과정에서 경험하지 않은 새로운 데이터 셋을 입력받아 이 데이터 셋에 최적화된 초기 가중치를 학습자에게 전달한다. 학습자는 이 가중치를 이용해 학습 샘플을 학습하여 Task T_{target}에 최적화된다. 이후 Task T_{target}의 테스트 샘플을 검증하여 최종적으로 메타 학습을 이용한 골 성숙도 분류 문제의 성능을 나타낼 수 있다.

3.2 실험 방법

제안하는 방법의 성능을 평가하기 위해 다음 두 데이터 셋에 대해 실험을 진행하였다. 첫 번째로, 메타 학습에 사용했던 메타 데이터 셋의 검증 샘플을 사용하여 성능을 평가한다. 메타 학습자 학습에 사용되었던 TinyImageNet과 Cifar100 데이터 셋 중 학습에 사용하지 않은 부류 중 5개를 임의로 선정하여 부류 당 20장의 영상을 사용해 학습한 후 15장의 검증 샘플로 성능을 평가한다. 따라서 총 75장의 샘플이 검증에 사용된다. 또한 동일한 학습샘플 100장과 검증 샘플 75장을 사용하여 일반적인 Base Learning 방법으로 모델을 학습한 결과와 성능을 비교한다.

두 번째로 실제 학습하고자 하는 수완부 영상 데이터를 사용하여 분류 정확도를 평가한다. 이 역시 앞선 방법과 마찬가지로 각 부류 당 학습 샘플 20장, 검증 샘플 15장을 이용해 메타 학습 방법과 Base Learning을 이용한 방법으로 모델을 각각 학습한다.

이때 각 데이터 셋에 대해서 두 가지의 학습 방법을 사용한다. 첫 번째로 컨볼루션 신경망에서 분류를 담당하는 완전 연결 층의 가중치 업데이트에만 메타 학습 방법을 적용한다. 이는 기존에 학습할 데이터량이 적을 때 사용하는 전이 학습(Transfer Learning)과 유사하다. 전이 학습은 사전 학습된 컨볼루션 층의 가중치는 고정하거나 가장 마지막의 컨볼루션 층만을 학습 과정에 포함하고 분류를 담당하는 완전 연결 층의 가중치를 실제 학습하고자 하는 데이터로

학습하여 적은 데이터만으로 과적합을 피하면서 학습할 수 있는 방법이다. 실험에서는 동일한 컨볼루션 층의 가중치를 가지면서 메타 학습으로 완전 연결 층의 가중치를 학습한 결과와 일반적인 전이 학습 방법으로 완전 연결 층의 가중치를 학습한 결과를 비교한다. 층과 가중치의 개수에 비해 좋은 성능을 내는 모델 구조인 VGG16[9]의 구조를 참고하여 신경망을 구성하였다. 신경망의 구조는 Fig. 4(a)와 같다. 깊이가 64, 128, 256, 512, 512인 컨볼루션 층이 각 2, 2, 3, 3, 3개로 총 13개의 컨볼루션 층과 256노드를 가는 완전 연결 층 3개로 구성된다. 사전에 학습된 컨볼루션 가중치는 대량의 공용 데이터 셋을 이용해 학습되어 온라인에서 제공되는 값을 사용하였다.

두 번째로 컨볼루션 층을 포함한 모든 가중치를 메타 학습의 가중치 업데이트 과정에 포함한다. 모든 네트워크의 가중치를 사전 학습 없이 초기 가중치로부터 학습하는 방법을 scratch 학습이라고 한다. 실험에서는 메타 학습으로 모든 가중치를 학습한 결과와 모든 가중치를 일반적인 Base Learning 으로 학습한 결과를 비교한다. 이 경우 MAML에서 사용했던 신경망 구조를 참고하여 3×3 크기의 컨볼루션 필터가 깊이 64개로 구성된 컨볼루션 블록 7개, 노드 256개인 완전 연결 층 3개로 구성된다. 신경망 구조는 Fig. 4(b)와 같다. 각 컨볼루션 블록은 모두 컨볼루션 연산, 배치 정규화, 활성화 함수 ReLU를 포함한다.

3.3 실험 데이터 구성

3.3.1 메타 데이터 셋 구성

Fig. 5는 메타 학습에 사용되는 학습 데이터 셋 중 n개의 과제가 포함된 하나의 배치(Batch)를 나타낸다. 딥러닝에서는 일반적으로 여러 개의 데이터를 하나로 묶은 배치 형태의 데이터를 학습하여 일반화 성능을 높인다. 메타 학습에서는 하나의 과제가 하나의 데이터이므로 여러 개의 과제가 묶인 형태의 배치를 학습한다. 학습에 사용되는 데이터는 공공데이터 셋으로 제공되는 TinyImageNet[6]과 Cifar100[7] 데이터 셋을 결합하여 총 300개의 부류를 가진 데이터 셋을 사용하였다. Cifar100은 주로 동물, 이동 수단, 물체 등의 영상으로 구성되어 있으며 각 부류는 600장의 영상을 포함한다. TinyImageNet은 그 외 풍경, 사람의 움직임 등의 복잡한 영상도 포함하며 한 부류당 500장의 영상으로 구성된다. 기존에 한 부류에서 학습용 샘플과 테스트용 샘플을 나누어 구성하던 데이터 셋과 달리 메타 학습은 여러 개의 부류로 이루어진 하나의 과제를 하나의 데이터로 학습하기 때문에 학습, 검증, 테스트에 속하는 데이터들은 모두 부류가 달라야 한다.

본 연구에서는 5개의 골 성숙도 부류를 분류하는 것이 목표이므로 메타 데이터 셋 역시 5개의 부류로 이루어진 학습 과제로 구성한다. 각 과제는 한 부류당 20개의 학습데이터, 15개의 테스트 데이터를 가지는 5개의 부류로 구성된다. 따라서 한 과제당 175개의 영상을 포함한다. 학습에는 100개의 샘플이, 테스트에는 75개의 샘플이 사용된다. 한 배치에는 2개의 과제가 포함되므로 한 배치에는 학습 샘플 200개, 테스트 샘플 150개로 총 350개의 데이터가 포함된다. 실험 과정에서는 300부류 중 임의로 10개의 부류를 선정하여 하나의 배치에 들어갈 2개의 과제를 구성했으며 이러한 배치 10000개를 생성하여 실험에 사용하였다.

3.2.2 수완부 영상 데이터 셋 구성

수완부 영상의 학습데이터는 5-18세 영유아 및 청

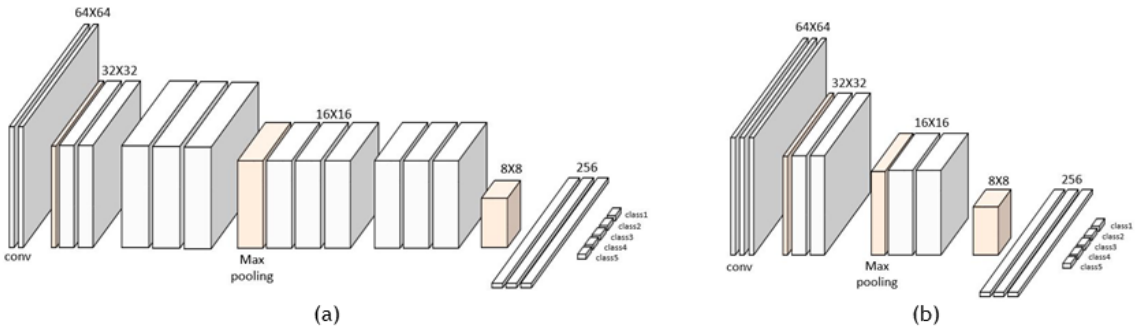

Fig. 4. Two CNN structures that, (a) learns only weight of FC (fully-connected) layers, (b) learn all layers.

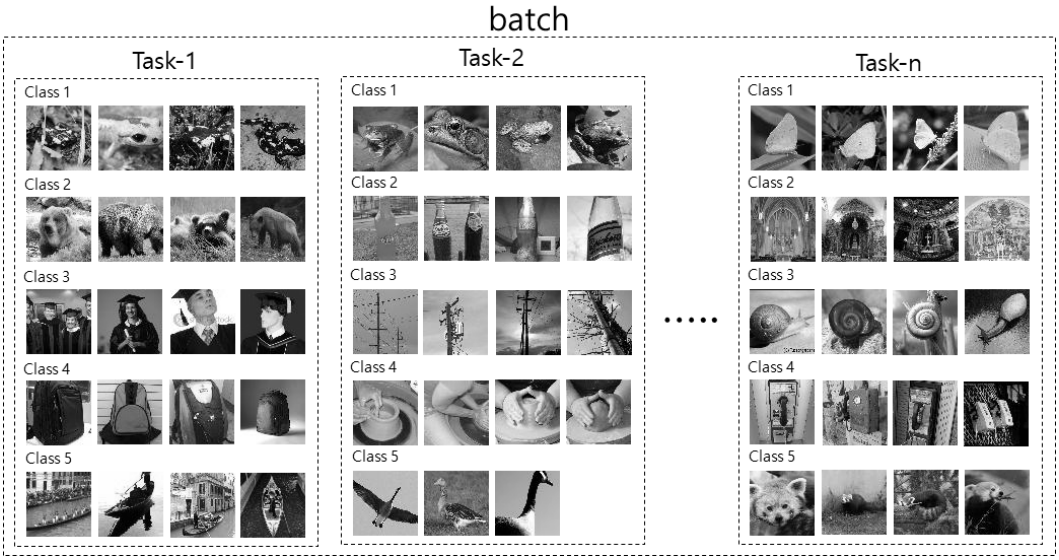

Fig. 5. Meta-Learning Dataset (TinyImagenet, Cifar100).

소년의 수완부 X-ray 영상 865장으로 구성된다. 성숙도는 총 5단계로 구분하며 각 단계 당 20장만의 영상을 학습에 사용하고 15장의 영상을 검증에 사용한다. 따라서 총 175장의 수완부 영상이 학습에 사용된다. 이때 전체 수완부 영상 보다는 각 손가락의 마디 같은 국소 영역이 골 성숙도 판별에 효율적이므로 512×512 크기의 전체 영상에서 6개의 ROI(Region of Interest)영역을 선정했다. ROI(Region of Interest)란 영상 내에서 집중하고자 하는 국소 영역을 의미한다. 영상 처리를 영상의 모든 영역에 대해 수행하게 되면 문제 해결에 중요한 영향을 미치는 정보가 작은 영역에 집중되어 있을 경우 모든 영역에 대해 연산을 하는 것이 비효율적이다. 따라서 작은 크기의 ROI(Region of Interest)를 지정하여 그 영역만 연산에 포함하는 방법을 사용한다. 이들은 골 성숙도 판단에 큰 영향을 미치는 엄지, 중지, 약지의 골단(epiphyseal) 영역이다. Fig. 6은 전체 수완부 영상 및 6개의 ROI를 나타내며 번호는 각 ROI의 순서를 의미한다. 각 ROI 영상은 64×64 크기이며 X-ray 영상이므로 한 채널만을 가진다.

ROI가 영상 당 6개 이므로 총 175×6=1050장의 64×64 크기 영상 데이터가 학습에 사용된다. Fig. 6은 잘라낸 ROI를 각 성장 단계별로 나타낸 것이다. 각 행은 6개의 ROI를 나타내며 열은 5단계로 분류할 부류(Class)를 나타낸다. 기존에 주어진 데이터는 골 성숙도를 총 11단계로 구분하며 한 단계에서 1-2년의 연령을 포함한다. 하지만 이 경우 한 부류에 속하는 영상이 최소 12장에서 최대 200장 이상으로 부류 간 데이터 량의 차이가 아주 심한 경우가 발생한다. 따라서 최소한의 학습데이터 확보를 위해 이웃한 두 부류끼리 데이터를 병합하여 5개의 부류로 성숙도를 구분한다. 병합 전과 후 부류 별 영상 수는 Table 1과 같다.

4. 실험 결과 및 고찰

4.1 메타 학습 결과

실험 결과는 아래와 같이 나타내었다. 부류 별 학습데이터를 5, 10, 20장으로 늘려가며 학습하였으며

Table 1. Number of data by class (before merging, after merging)

Before merging	Class(age)	1	2	3	4	5	6	7	8	9	10	11
	Number of data	188	145	277	68	10	48	75	2	11	29	12
After merging	Class(age)	1		2		3		4			5	
	Number of data	333		345		58		88			41	

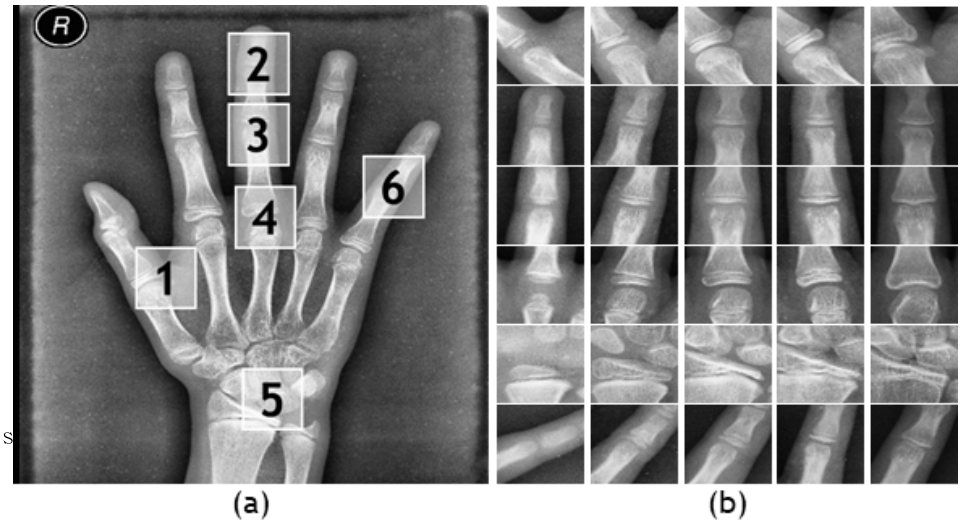

Fig. 6. Six ROIs(Regions of Interest) for skeletal maturity indication. (a) Hand-arm X-ray image with 6 ROIs, (b) Examples of segmented ROIs.

테스트 샘플은 모든 경우 부류 당 15장으로 고정하였다. 본 실험은 학습에 사용되는 데이터 개수가 적기 때문에 데이터 셋이 어떻게 구성되느냐에 따라 성능 편차가 존재한다. 따라서 같은 실험을 5 번씩 반복하여 평균 성능을 나타낸다. 성능 지표인 정확도는 검증에 사용한 75장의 데이터 중 자신의 부류에 옳게 분류 된 데이터 수의 비를 퍼센트로 나타내었다. 기존의 학습 방법인 Base Learning과 메타 학습 방법을 비교하였으며 각각의 방법들은 사전에 학습된 컨볼루션 층의 가중치를 사용한 Pre-trained 방법과 임의로 초기화된 가중치를 처음부터 학습하는 Scratch 방법 두 가지를 사용하였다. 또한 결과는 메타 검증 데이터 셋과 실제 분류하고자 하는 수완부 영상 데이터 셋을 사용하였다. 메타 검증 데이터 셋을 사용한 결과는 Table 2에 나타내었다. Pre-trained, Scratch 학습 모두 메타 학습을 이용한 학습결과의 정확도가 Base Learning에 비해 높은 성능을 보였다. 메타 검증 데이터 셋의 샘플들은 메타 학습과정에 사용된 부류들과 동일하진 않지만 같은 데이터 셋 안에 포함된 다양한 부류들로 구성되어 있어 성능이 비교적 높은 것으로 보인다.

Table 3은 수완부 영상을 이용한 학습 결과를 나타낸다. 메타 학습 결과 학습 데이터 부류 별 20장을 사용하여 Pre-trained 방법으로 학습한 결과 정확도 69.3%, Scratch 학습 결과 정확도 65.9%로 모두 비교 실험인 Base-Learning 방법에 비해 높은 성능을 보였다. 비교 실험의 경우 컨볼루션 층의 가중치를 모두 기존 방법으로 학습한 Scratch 학습의 경우는 학습이 잘 진행되지 않았으며 특히 5, 10장을 이용한 학습은 임의로 5개의 부류 중 한 가지를 선택할 때의 확률인 20%에 가까운 결과를 보였다. 이는 특징 추출 과정에서 데이터가 매우 부족했음을 의미한다.

또한 6개의 ROI 별 분류 성능 차이도 존재하며 이는 Table 4와 같다. ROI 2는 중지 끝마디를, ROI

Table 2. The Accuracies for proposed meta-learning method and conventional base learning method

Meta Dataset Learning Accuracy(%)					
Method		Base Learning		Meta Learning	
Model		Pre-trained	Scratched	Pre-trained	Scratched
number of sample	5	51.9	27.3	73.3	53.3
	10	58.6	31.2	76.0	59.9
	20	69.3	35.9	**81.9**	**80.0**

Table 3. The Accuracies for proposed meta-learning method and conventional base learning method

SMI Dataset Learning Accuracy(%)					
Method		Base Learning		Meta Learning	
Model		Pre-trained	Scratched	Pre-trained	Scratched
number of sample	5	42.4	23.1	47.1	53.3
	10	45.7	31.4	57.1	56.6
	20	50.6	38.6	**69.3**	**65.9**

4는 중지 첫마디에 해당하는 영역이다. Fig. 7에서 중지 첫마디는 연령이 올라갈수록 마디의 간격이 좁아지는 반면 끝마디는 연령 별로 마디의 간격이나 너비에 변화가 적다. 이러한 이유로 인해 영역 별로 분류 성능에 편차가 존재하므로 실험에서는 각 영역 별로 최대값을 가지는 부류를 봇팅(votting)하여 가장 많이 봇팅된 부류를 최종 부류로 선정한다. 이때 각 영역 별로 가중을 동일하게 부여하여 봇팅하는데 실제 영역 별로 분류 결과에 성능 차이가 있으므로 가중을 다르게 부여한다면 최종 분류 결과가 더 정확할 것이다. 특히 이 가중치를 신경망 학습과정에 포함하여 같이 학습한다면 성능을 개선할 수 있을 것이다. 하지만 MAML 방법을 사용한 메타 학습 과정에서 이러한 가중치를 학습하려면 그 신경망의 구조가 메타 학습 과정에도 포함되어야 한다. 따라서 메타 학습 과정에서도 마찬가지로 임의의 ROI를 지정하여 최종 출력의 크기를 수완부 영상을 이용한 학습의 테스트 과정과 동일하게 구성해야 하는 어려움이 있다.

5. 결 론

본 논문에서는 딥 러닝 기반의 메타 학습을 이용해 적은 양의 학습 데이터만으로 수완부 영상에서 골 성숙도를 분류한다. 메타 모델은 적은 수의 샘플만이 포함된 데이터 셋을 여러 개 이용하여 학습을 반복하며 손실에 민감한 가중치를 갖도록 훈련된다. 비교실험을 위해 동일한 100장의 샘플을 사용해 일반적인 전이학습과 성능을 비교하였으며 메타 학습이 더 높은 성능을 보임을 확인하였다. 먼저 메타 학습의 한 방법인 MAML(Model-Agnostic Meta Learning) 방법을 이용하여 다양한 학습과제로 이루어진 별도의 데이터를 사용하여 메타 모델을 학습한다. 그 후 학습된 메타 모델을 사용하여 수완부 X-ray 영상을 성숙 단계별로 분류한다. 이때 수완부 영상은 6개의 ROI를 지정하여 각 영역에 대한 분류 결과를 봇팅(votting)하여 최종 성능을 나타낸다. 이때 특징 추출을 위해 사전에 학습된 VGG16 모델의 컨볼루션

Table 4. Classification Accuracy of each ROI (Region of Interest)

ROI	1	2	3	4	5	6
Accuracy(%)	63.0	45.9	50.0	69.9	59.9	50.0

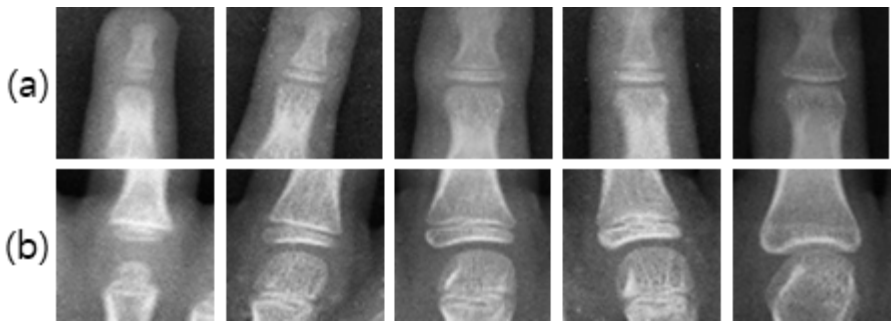

Fig. 7. Skeletal change according to each age group (a) ROI 2(Middle finger last node), (b) ROI 4(Middle finger first node). The right side shows higher ages.

가중치를 사용한 Pre-trained 방법과 초기화된 가중치를 처음부터 학습하는 Scratch 방법을 이용해 그 성능을 비교한다. 한편 본 논문에서는 골단에 해당하는 6개의 영역을 ROI(Region of Interest)로 지정하였으나 골 성숙도 판단에는 더 다양한 영역이 사용된다. 이 영역들을 ROI에 반영하여 학습 데이터 셋을 구성한다면 더 좋은 성능을 기대할 수 있을 것이다. 봇팅 과정에서는 각 영역 별 출력에 대한 가중을 동일하게 주었다. 하지만 사용하는 ROI의 개수를 늘리고 상대적으로 더 중요한 영역에 가중치를 크게 준다면 성능 개선을 기대해 볼 수 있다. 또한 메타 모델의 사전 훈련에 사용되었던 TinyImageNet과 Cifar100은 다양한 이미지를 포함하지만 수완부 영상의 경우 형태가 유사한 여러 영상을 분류해야 한다. 그러므로 메타 모델의 사전 훈련에 부류별 이미지 형태가 유사한 데이터를 다량 포함하여 학습을 진행한다면 조금 더 좋은 성능을 기대할 수 있을 것이다. 마지막으로, 본 논문에서 사용한 데이터 셋은 성별이 분류되어 있지 않다. 골 성숙도는 성별에 따라 같은 연령이라도 많은 차이가 나기 때문에 성별은 골 성숙도 판별에 중요한 역할을 한다. 성별에 대한 정보를 포함하는 양질의 데이터 셋이 제공된다면 성별을 분류하는 네트워크를 추가로 학습하여 좀 더 좋은 성능을 기대할 수 있을 것이다.

REFERENCE

[1] Yunju Jeong, Israfil Ansari, Jaechang Shim and Jeonghwan Lee, “A Car Plate Area Detection System Using Deep Convolution Neural Network“, *Journal of Korea Multimedia Society,* Vol. 20, No. 8, pp. 1166-1174, August 2017

[2] S. Ravi and H. Larochellem, “Optimization as a Model for Few-Shot Learning,” *Proceeding of International Conference on Learning Representations*, Vol. 1, pp. 6, 2017.

[3] C. Finn, P. Abbeel, and S. Levine, “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks,“ *Proceeding of International Conference on Machine Learning*, Vol. 70, pp. 1126-1135, 2017.

[4] Z. Li, F. Zhou, F. Chen, and H. Li, “Meta-SGD: Learning to Learn Quickly for Few-Shot Learning,“ arXiv preprint arXiv:1707.09835, 2017.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based Learning Applied to Document Recognition,” *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, 1998.

[6] A. Bastidas, Tiny ImageNet Image Classification, https://tiny-imagenet.herokuapp.com/ (accessed Jan., 12, 2018)

[7] A. Krizhevsky, Learning Multiple Layers of Features from Tiny Images, https://www.cs.toronto.edu/~kriz/cifar.html (accessed Jan., 12, 2018).

[8] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *Proceeding of International Conference on Machine Learning*, Vol. 37, pp. 448-456, 2015.

[9] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,“ arXiv preprint arXiv:1409.1556, 2014.

민 정 원

2016년 부산대학교 기계공학부 졸업
2016년~2017년 부산대학교 대학원 기계공학부 석사
2018년~현재 LG전자 근무 중
관심분야 : 머신비전, 딥 러닝

강 동 중

1988년 부산대학교 정밀공학과 졸업
1990년 한국과학기술원 기계공학 석사 졸업
1999년 한국과학기술원 자동화 및 설계공학 박사 졸업
1997년~2000년 삼성 종합기술원 신호처리연구실.
2006년~현재 부산대학교 기계공학부 제어자동화시스템 전공 교수
관심분야 : 머신 비전, 패턴 인식, 공정 검사 자동화