

## Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

### 1. Abstract

deep neural network를 학습하는 과정은 그 전 layer의 parameter가 변경됨에 따라 훈련(train) 중에 각 layer의 input이 바뀌기 때문에 굉장히 복잡하다. 이는 낮은 learning rate와 parameter 초기화를 요구하여 nonlinear한 model을 train하기 어렵게 만든다.

우리는 이 현상을 'internal covariate shift' 라고 하며, layer의 input을 normalize해서 문제를 해결한다. 우리의 method는 normalization을 모델 아키텍처의 일부로 만들고 각 미니 배치에 대해 normalization을 수행하는 것에서 강점을 얻는다. 또한 regularizer 역할을 하기 때문에 어떤 경우에는 dropout이 필요하지 않다. 최첨단 이미지 분류 모델에 적용된 batch normalization은 14배 적은 training 단계로 같은 accuracy를 보이고 원래 모델을 상당한 차이로 능가했다. 그리고 batch normalized된 network의 앙상블을 사용해 ImageNet 분류에 대해 4.9% top-5 validation error(4.8%의 test error)를 달성해 인간의 정확도를 능가하였다.

### 2. Introduction

SGD(Stochastic gradient descent)는 training deep networks의 효율을 높여주었다. 하지만 세심한 hyper parameter 튜닝을 필요로 한다. parameter의 작은 변화들이 모델에 큰 영향을 미치기 때문이다.

$$\Theta = \arg \min_{\Theta} \frac{1}{N} \sum_{i=1}^N \ell(x_i, \Theta)$$

이것은 SGD를 계산하는 식이다. SGD에서 각 step은 사이즈가 m인 mini-batch라고 생각할 수 있다. mini-batch는 각각의 parameter에 대한 loss function의 gradient를 근사(approximate)하는 데에 사용된다.

이전까지는 Gradienet Vanishing, saturation problem의 해결방법으로 ReLU와 small learning rates 방법이 제안되어왔다. 이 논문에서는 다른 해결방법으로 batch Normalization을 제안한다. batch norm은 layer input data의 평균과 분산을 고정시켜 데이터가 일정한 분포를 가지게 한다. 또한 parameter의 초기값과 크기에 대한 gradient의 의존도를 줄여 gradient flow에 긍정적인 영향을 준다. 이로 인해 더 높은 learning rate 사용이 가능해지고 dropout의 필요성이 줄어든다.

### 3. batch Normalization

<b>Input:</b> Values of $x$ over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$ ; Parameters to be learned: $\gamma, \beta$	
<b>Output:</b> $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$	
$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$	// mini-batch mean
$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2$	// mini-batch variance
$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$	// normalize
$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)$	// scale and shift

**Algorithm 1:** Batch Normalizing Transform, applied to activation  $x$  over a mini-batch.

이것은 batch-normalization을 계산하는 과정이다. 감마는 scaling, 베타는 shift, 입실론은 연산 수행시 안정성을 위해 추가하는 아주 작은 크기의 상수이다.

입력 데이터가 정규화되면 대부분의 입력에 대해 매우 선형적으로 동작하기 때문에 non-linearity를 유지하기 위해서 scaling, shift과정을 거친다.

## YOLO 기반의 광학 음악 인식 기술 및 가상현실 콘텐츠 제작 방법

### 1. 서론

OMR(Oprical Music Recognition)은 사진 형태의 악보를 디지털화시키는 것을 말한다. 이 연구를 통해 음악 표기가 모호하거나 질이 좋지 않은 악보를 개선하여 다양하게 활용할 수 있게 된다. 또한 OMR은 컴퓨터 비전, 디지털 음악학, 디지털 라이브러리 등 여러 분야와 광범위하게 관련되어 있기 때문에 딥러닝의 발전에 좋은 영향을 줄 수 있다.

### 2. 본론

먼저 악보 위의 음표를 해석하기 위해서는 1) 음표의 종류를 파악하고, 2) 오선의 몇 번째 줄에 위치하는지 파악 3) 음표의 Beam, Flag의 존재 여부를 파악해야 한다. 따라서 YOLO를 활용해 epoch 1000, batch size 2, image size 2048x2048x3으로 학습을 진행하였다. 그리고 학습을 마친 뒤에는 학습한 YOLO 모델 및 CV 알고리즘을 활용하여 악보의 요소를 탐지하였다. 주로 음표, 쉼표, 음자리표 등이다.

### 3. VR 콘텐츠 생성

악보 검출로 얻은 데이터는 콘텐츠 진행의 핵심 요소인 '채보'로 취급된다. 이 채보에는 음정,박자 등 곡의 전반적인 요소가 들어 있다. 이 게임이 정상적으로 실행되기 위해서는 채보 파일과 재생될 음악 파일, 음악의 선택을 도와줄 이미지 파일이 필요하다. 이 3종류의 파일을 모두 디렉토리에 추가하고 언리얼에서 이벤트가 발생했을 때, 디렉토리 안에 존재하는 모든 노래에 대한 정보를 배열에 넣어 동기화한다. 동기화는 두 가지 상황에서 일어나는데, 먼저 어떤 레벨에서 곡 선택 레벨로 전환되었을 경우와 게임을 본격적으로 시작하는 메인 레벨로 이동했을 경우이다.

### 4. 결론

이 논문에서는 YOLO v5를 사용해 악보에 대한 OMR 작업을 진행하였고 이를 바탕으로 VR 콘텐츠 제공을 제안하였다. 이 OMR 과정에서는 악보의 기호들이 다양한 모양을 가지고 있었기 때문에 Hough transform을 이용해 검출하였다.