Speech Recognition

음성인식 실습

경북대학교 인공지능학과 김준우 박사과정
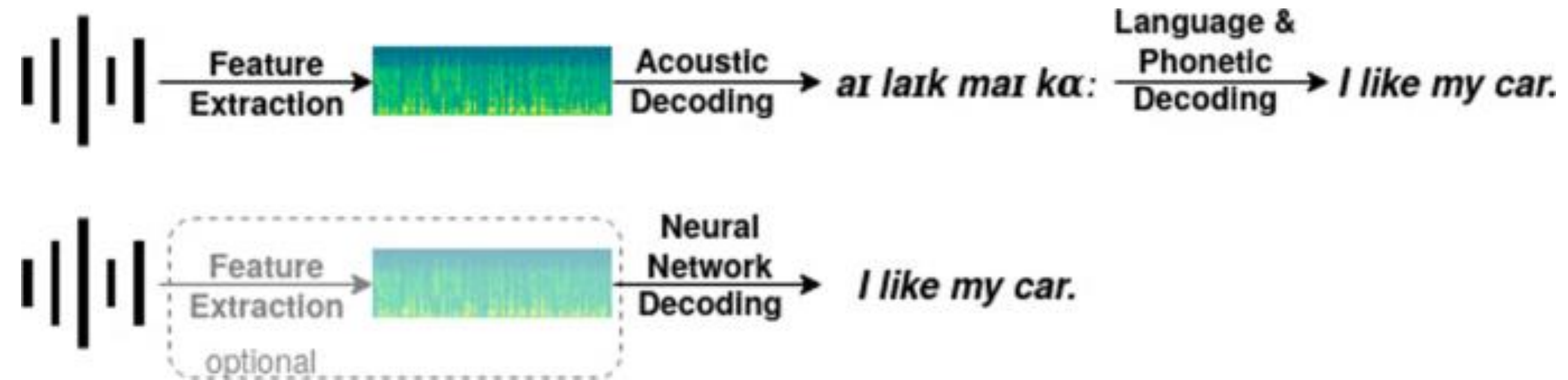
kaen2891@gmail.com

15, Feb. 2022.

MLCL Kyungpook National University

Deep Speech2 Model

End-to-End ASR Model

- End-to-end ASR model
  - Pipeline (top) vs end-to-end (bottom) ASR

Kyungpook National University
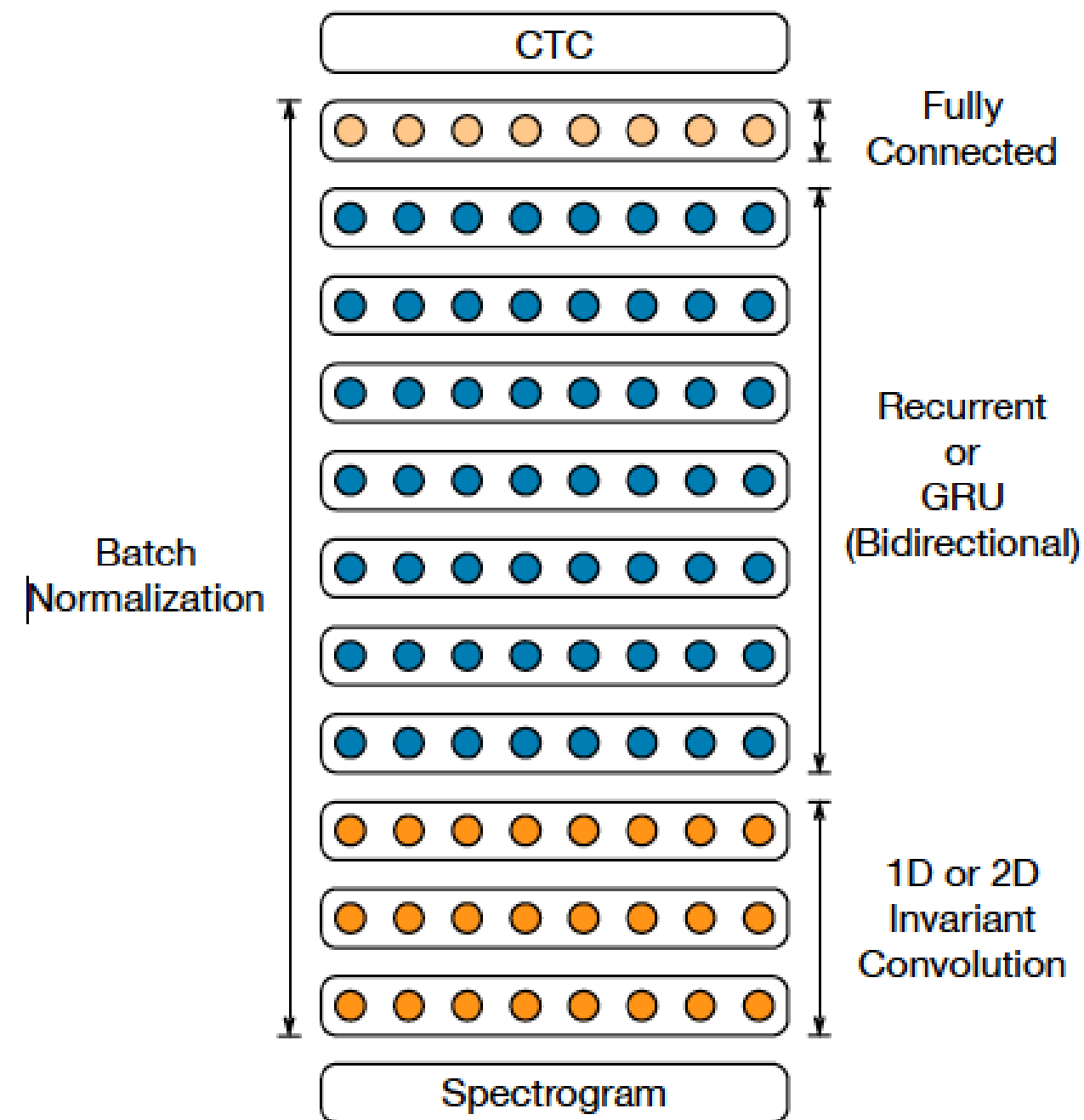
End-to-End ASR Model

- End-to-end ASR model: Deep Speech2
  - Model architecture

audio   spectrogram $P_t(c)$ training

Preprocessing → **DNN** → **CTC Loss**

- Data augmentation
  - additive Gaussian noise
  - time stretch (resampling)
- Windowing 20-25 ms, stride 10 ms
- FFT, log
- Normalization

- 2-3 convolutional layers
  - ch=32, ks=[11, 41], s=[2, 2]
  - ch=32, ks=[11, 21], s=[1, 2]
  - ch=64, ks=[11, 21], s=[1, 2]
- 3-7 bi-/uni- directional GRUs/LSTMs
- 1 row conv layer (for unidirectional)
- 1-2 fully connected layers

- BN/dropout for regularization

inference

Decoder

- greedy (argmax)
- beam search
- beam search with language model (width=2000-8000)

Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in english and mandarin." *International conference on machine learning.* PMLR, 2016.

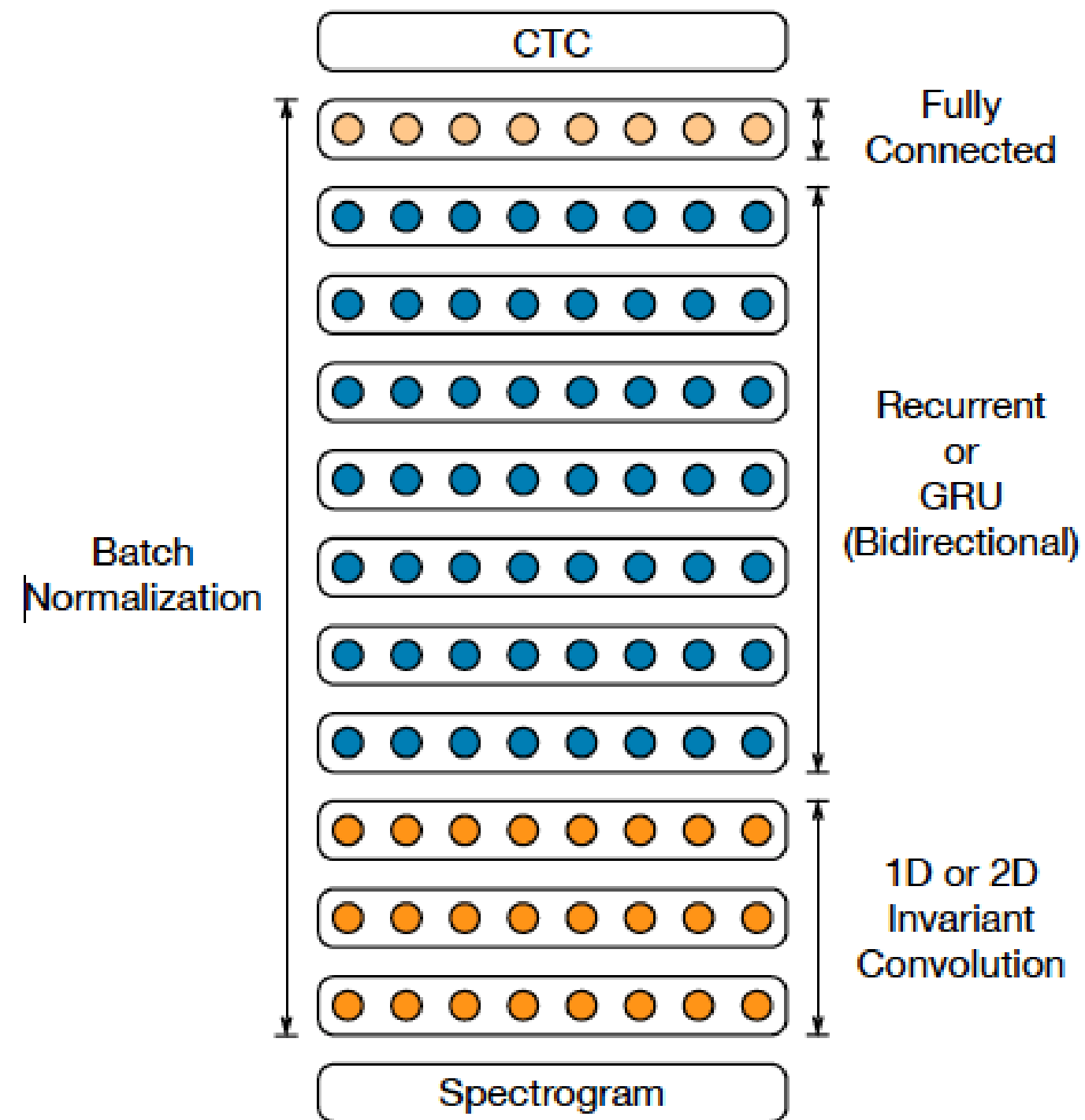MLCL **Kyungpook National University**

- End-to-end ASR model: Deep Speech2
  - Model architecture

Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in english and mandarin." *International conference on machine learning*. PMLR, 2016.
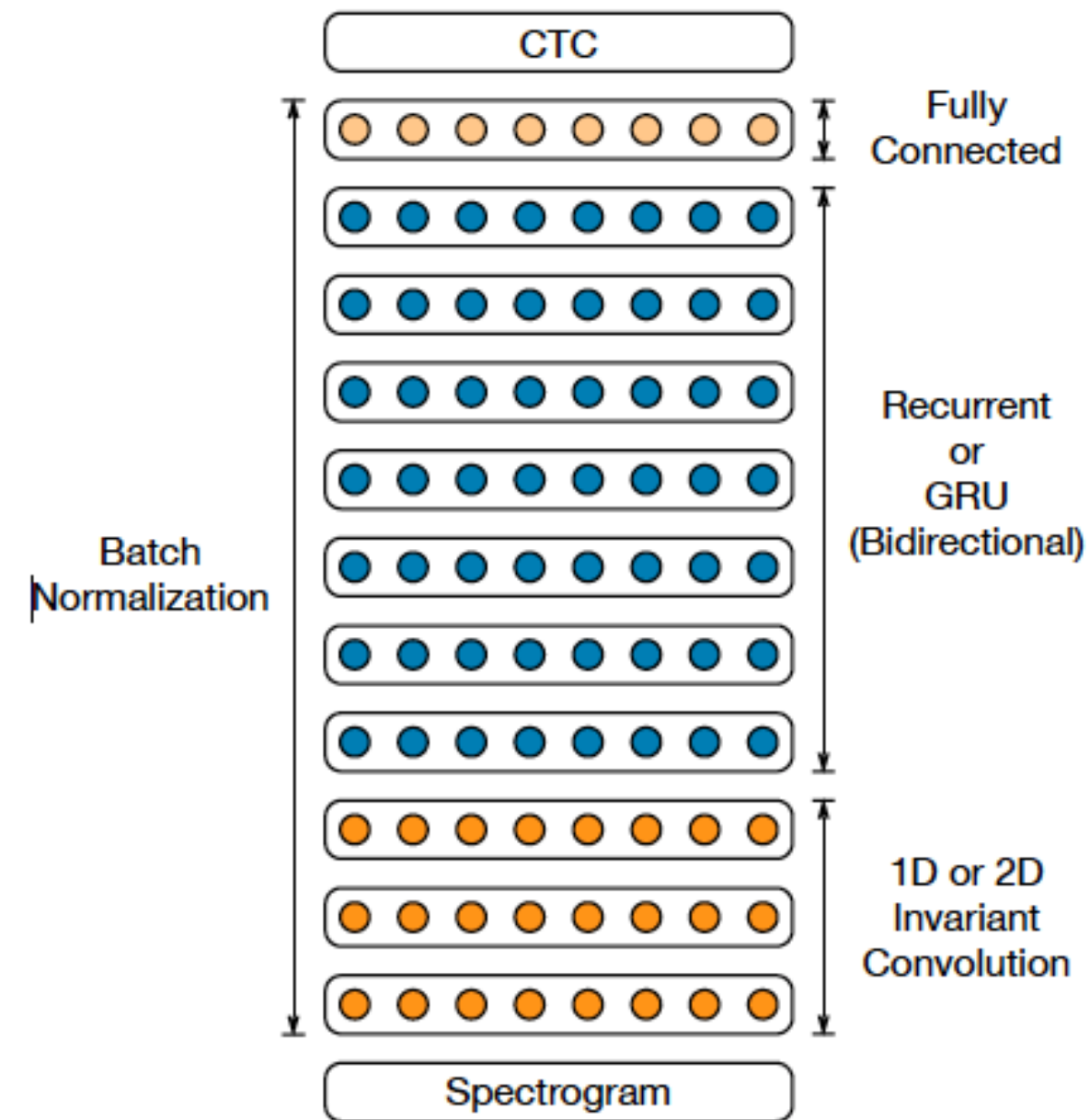
Deep Speech2 Model

- Composed of 11 layers
  - 3-layer of 2D-CNNs
  - 7-layer of RNNs
  - 1-layer of output (softmax)
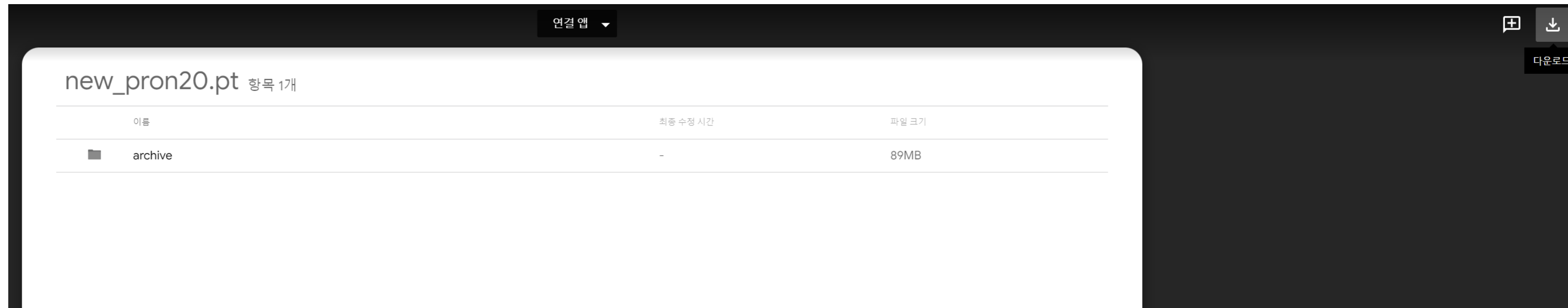 - Batch Normalization
  - ReLU nonlinearities

Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in english and mandarin." *International conference on machine learning*. PMLR, 2016.

Kyungpook National University

Deep Speech2 Model

- Colab 실습
  - https://colab.research.google.com/drive/14CexTcSHh-dZ1pBkljZbYZUAqF2sCugu?usp=sharing

Kyungpook National University

Pretrained weight

- Pretrained weight
  - 학습이 오래 걸리기 때문에, 미리 학습해 놓은 20에폭 모델 다운받기
  - https://kaen2891.tistory.com/122

  - 클릭하여 다운

Kyungpook National University

Pretrained weight

- Pretrained weight
  - 다운받은 new_pron20.pt를 colab 파일내에 드래그하여 이동

Kyungpook National University
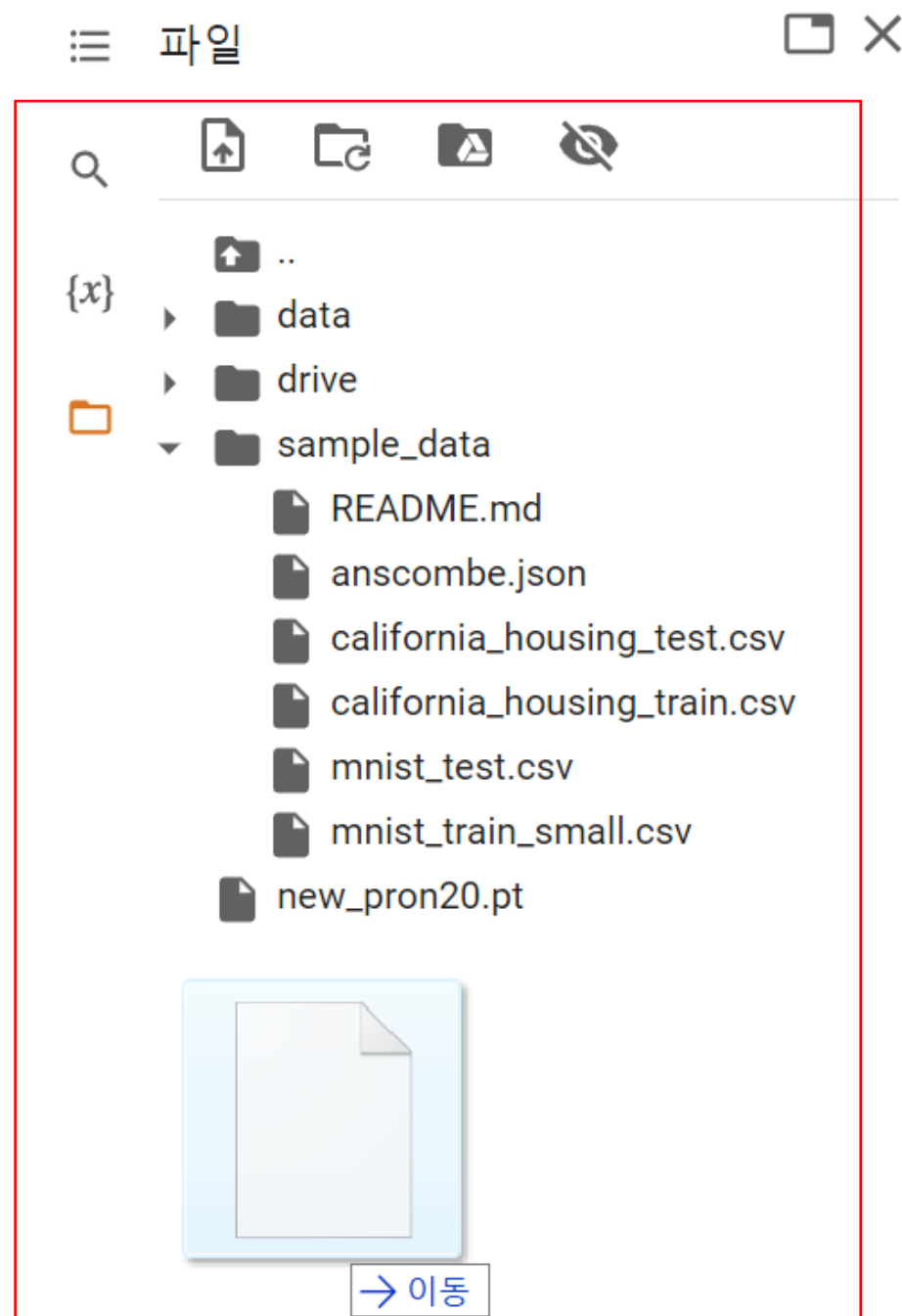
Thank you!

Presentation finished

MLCL

Kyungpook National University