

東工大TSUBAMEの利用事例 マルチメディア認識のための深層学習

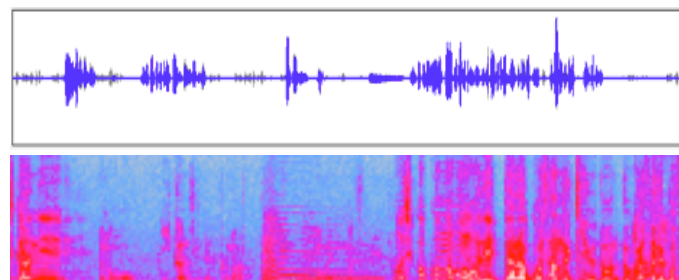
井上 中順, 篠田 浩一
東京工業大学 情報理工学院
E-mail: {inoue, shinoda}@c.titech.ac.jp

マルチメディア認識と深層学習

- 深層学習: 多層ニューラルネットワークの学習
 - 従来の手法に比べ**圧倒的に高性能**
 - ビッグデータの処理に**大規模計算機が必須**
 - **GPU**との親和性が高いアルゴリズムが多数
 - マルチメディア認識
 - 数千万のパラメータを**数百時間**かけて学習
 - 音声認識: DNN-HMM
 - 映像認識: CNN
- TSUBAME利用
研究例を紹介

音声認識

- Deep Neural Network + Hidden Markov Model
深層学習を用いた音響モデル, 約3,000万パラメータ
- **Student-Teacher型学習** [Price 2016]
 - 大型DNNを小型化する半教師付き学習
 - 304万パラメータ (90%圧縮)
 - 1,100時間のデータを学習
 - Yahoo!JAPANとの共同研究



R. Price, K. Iso, and K. Shinoda, "Wise Teachers Train Better DNN Acoustic Models," EURASIP Journal on Audio Speech and Music Processing, vol.10, pp.1-19, 2016.

映像認識

- Convolutional Neural Network

画像認識で広く用いられる, 約1,000万パラメータ

- **映像からの物体・動作検出** [Yamamoto 2015]

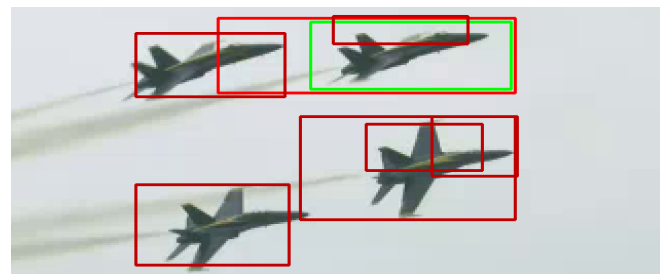
- CNNを映像認識に応用

- 1,000万フレームの処理 \approx 約500時間

- TRECVID 2015国際ワークショップでの評価

検索精度: 20チーム中3位

位置検出精度: 6チーム中3位



R. Yamamoto, N. Inoue, and K. Shinoda, "Localization with Spatio-Temporal Selective Search and SPPnet (TokyoTech at TRECVID 2015)," Proc. TRECVID, 2015.

TSUBAME3.0に向けて

- **Co-design**型の研究開発 (JST-CREST)

- 目的: 高精細映像のリアルタイム処理
- システムからアプリケーションまで全体最適化
- 分野を横断した研究開発体制

e.g., 深層学習ネットワーク構造とGPU間通信の最適化

- 複数GPUを活用した**大規模学習**

- 学習の並列化による高速化
- 従来よりはるかに大規模なデータ・モデルを扱える