The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

多人数追跡のための分散深層学習による高精度な検出にむけて

Towards Accurate Multiple Human Tracking Using Scalable Distributed Deep Learning

佐藤　仁 *1　　西川　由理 *1*2　　小澤　順 *1
Hitoshi Sato　　Yuri Nishikawa　　Jun Ozawa

*1産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

*2パナソニック
Panasonic

Scalable distributed deep learning is widely studies for open datasets for visual object recognition, i.e., ImageNet. In general, when we apply these scalable techniques to real applications, costly model training to domain-specific datasets is required for accurate recoginition; however, there are a few case studies for distributed deep learning except for ImageNet in terms of scalability, hyper parameter settings, and generalization, etc. This paper demonstrates our early activities on accurate human detection from soccer video images using distributed deep learning, as an instance towards accurate multiple human tracking from application-domain-specific video images.

1. はじめに

ビデオカメラで撮影された動画に対する多人数追跡は，小売店，物流センター，工場などでのヒト・モノの流れの分析・解析など幅広い応用が見込まれるため重要な技術である．とりわけ，サッカー，バスケットボール，アメリカンフットボールなどのチームスポーツの動画の分析・解析は，人物の動きが速い一方で，局所的な空間に人物が集中しオクリュージョンが頻発するため，高度な多人数追跡技術を必要とする事例の一つである．我々は，これまで，K-Shortest Path [Berclaz 11] のグラフ最適化アルゴリズムを用いた多人数追跡システムのスポーツ動画への応用を進めており，1 時間程度のバスケットボール動画の追跡においても現実的な時間で処理が可能であるという知見を得ている [Nishikawa 17].

多人数追跡は，一般に，(a) 動画のフレーム毎に人物を検出する，(b) 各フレームの人物検出結果を対応付けして軌跡を生成する，という 2 つのステップから成る．現状の我々のシステム [Nishikawa 17] では，(a) のステップにおいてロバストな人物検出を行うために，深層学習ベースの一般物体検出アルゴリズムである YOLO [Redmon 17] を用いている．しかし，人物検出の際の学習モデルは一般物体検出を目的とした COCO データセットより生成されたものを用いているため汎用の人物検出に特化しており，実問題を対象とした高精度な物体検出を行うためには，そのドメインに特化したデータセットより学習モデルを生成することが必要である．一方で，学習モデルを生成するためには大量の計算機資源を必要とし，その処理にも長時間を要することが問題となる．

分散深層学習は，複数の計算機を利用し並列分散処理により深層学習を行うことで，高速に学習モデルを生成することができると期待されている．特に，近年，ImageNet のデータセットに対する一般画像認識を対象に，大規模並列分散環境下での分散深層学習の高速化の取り組みが進んでいる．しかし，ImageNet 以外のデータセットを対象とし分散深層学習を用いた場合，学習率やバッチサイズなどのハイパーパラメタや汎化性能，最適化アルゴリズム，並列化性能などを評価した事例は多くはない．

連絡先: 佐藤　仁，産業技術総合研究所，東京都江東区青海 2－4－7　産業技術総合研究所臨海副都心センター別館，hitoshi.sato@aist.go.jp

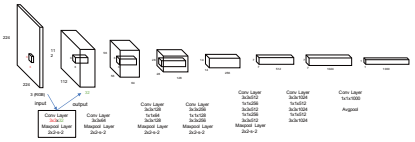

図 1: Darknet19

我々は，ドメインに特化したデータセットに対して分散深層学習により高精度な学習モデルの生成を高速に行うことで，多人数追跡の実問題への適用を目指している．本稿では，サッカー動画のデータを用いた多人数追跡を対象とし，この取り組みの概要と初期の評価について報告する．

2. ドメインに特化した多人数追跡

2.1 データセット

ドメインに特化した多人数追跡として，ここでは，サッカー動画のデータを用いた多人数追跡を対象とし，その際の人物検出の学習モデルを生成する例を取り上げる．ISSIA CNR では，イタリアのサッカー動画と人物を矩形としてアノテーションしたデータを公開している [D'Orazio 09]．これらの動画を基にして，人物を検出するための学習モデルを生成する．具体的には，動画のデータセット中から 5997 枚を画像として切り出し，それらの画像から総計 29860 個の人物の矩形情報を切り出すことで学習データセットとした．実際には，23836 個の人物の矩形情報を学習対象のデータセット，6024 個の人物の矩形情報をテスト対象のデータセットとした．

2.2 学習モデル

サッカー動画の人物検知のための学習モデルとしては，一般物体検出アルゴリズム YOLO で用いられている Darknet19 のネットワークを用いた [Redmon 17]．図 1 にそのネットワークの概要を記す．

3. 分散深層学習

3.1 分散深層学習フレームワーク

2.2 節の学習モデルの生成を分散深層学習で行うためのフレームワークとして ChainerMN [Akiba 17] を用い，データセットを分散して学習モデルを計算する「データ並列」の手

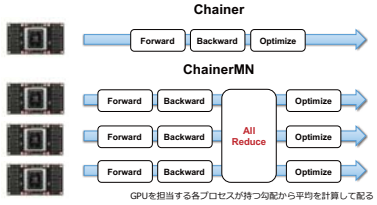

図 2: ChainerMN による分散深層学習

表 1: AAIC の計算ノードのスペック

CPU	Intel Xeon E5-2630L v4 1.80GHz (10 cores, HT-enabled) × 2
GPU	NVIDIA Tesla P100 × 8
Mem	256 GiB
SSD	Intel DC S3510 480 GB × 1

法を用いた．その概要を図 2 に記す．深層学習では，1) 予測を行いその誤差を計算（Forward 処理）し，2) 誤差を減らす方向の勾配を計算（Backward 処理）し，3) 勾配を用いて学習モデルを更新する．ChainerMN による分散深層学習では，1) Forward 処理，2) Backward 処理を行った後，分散して計算した勾配から平均を計算し配り直す AllReduce 処理を行う．そのため，計算性能だけでなくネットワーク通信の性能が重要となる．

3.2 ハイパーパラメタの設定

分散深層学習は精度を保って高速化することが難しい．具体的には，a) どのような勾配降下の最適化アルゴリズムを用いるか，b) どのような学習率やデータセットのバッチサイズなどのハイパーパラメタを設定するか，c) いかに局所的な最適解を避け，大域的な最適化を行うか，などが課題となる．今回は，[Goyal 17] の論文と同様の方法を用いた．具体的には，バッチサイズを GPU 台数 ×32 (1GPU 毎に 32)，学習率を 0.1× バッチサイズ /256，最適化アルゴリズムを Momentum SGD (momentum=0.9)，Weight decay を 0.0001 とした．実験は epoch 数を 10 として行った．

4. 初期評価

4.1 実験環境

初期の評価を産総研 AI クラウド AAIC 上で行った．表 1 に計算ノード 1 台のスペックを示す．計算ノード間は EDR Infiniband により Director Switch を介して Full-bisection Fat Tree 構成で接続されている．計算ノードの OS は CentOS 7.3 で構成され，Linux のカーネルは v3.10.0 である．また，ソフトウェアは，GCC v4.8.5, CUDA v8.0.61.2, CuDNN v6.0.21, NCCL v2.1.4, MPI は OpenMPI v2.1.2 をベースとして，ChainerMN v1.2.0, Chainer v3.4.0, CuPy v2.4.0 を用いた．

4.2 実験結果

図 3 に結果を記す．x 軸は GPU の台数，y 軸は 1 台の GPU を用いた際の実行時間を基準としたときのスケーラビリティを表す．概ねスケーラブルな処理性能を示すことを確認し，16 台の GPU で 6.41 倍の性能を達成した．

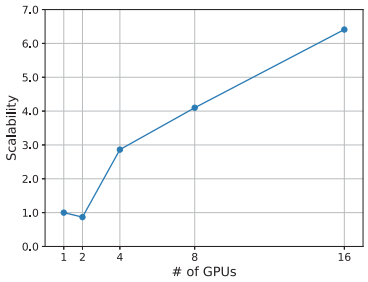

図 3: 実験結果

5. おわりに

ドメインに特化したデータセットに対して分散深層学習により高精度な学習モデルの生成を高速に行うこと目指し，サッカー動画のデータを用いた多人数追跡を対象とした事例について報告した．今後の課題としては，数千 GPU 規模の環境での大規模な分散深層学習の実行に向けて，a) ドメインに特化した大規模データセットの収集方法やラベル付け，b) 大規模なバッチサイズ下でのハイパーパラメタの設定，c) ネットワーク通信や I/O の最適化，などの検討を進めていきたい．

参考文献

[Akiba 17] Akiba, T., Fukuda, K., and Suzuki, S.: ChainerMN: Scalable Distributed Deep Learning Framework, in *arXiv.org e-Print archive*, pp. 1–6 (2017)

[Berclaz 11] Berclaz, J., Fleuret, F., Turetken, E., and Fua, P.: Multiple Object Tracking Using K-Shortest Paths Optimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 9, pp. 1806–1819 (2011)

[D'Orazio 09] D'Orazio, T., Leo, M., Mosca, N., Spagnolo, P., and Mazzeo, P.: A Semi-automatic System for Ground Truth Generation of Soccer Video Sequences, in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 559–564 (2009)

[Goyal 17] Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Kaiming, Y. J., and Facebook, H.: Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour, in *arXiv.org e-Print archive*, pp. 1–12 (2017)

[Nishikawa 17] Nishikawa, Y., Sato, H., and Ozawa, J.: Performance evaluation of multiple sports player tracking system based on graph optimization, in *The IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD 2017) in IEEE BigData2017*, pp. 2903–2910 (2017)

[Redmon 17] Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525 (2017)