

バッチを用いたディープラーニング特徴抽出

Context

Introduction

現代の深層学習モデルは、同じミニバッチ内のサンプル間の相互作用（バッチレベル相互作用）を活用することで、画像のより強力な特徴表現を学習してきました。特に、画像分類における表現学習の最近の進展（特に2020年以降）は、バッチ内の他のサンプル（否定例、肯定例、または文脈参照として）を考慮することで、学習された特徴が大幅に改善されることを示しています。本報告書では、このようなバッチコンテキスト技術を活用する主要な研究論文（2020年以降）を調査します。各論文については、タイトル、著者、年、方法の主要なアイデアに焦点を当てた要約、バッチレベル相互作用の活用方法、および論文へのリンクを一覧で示します。論文は、アプローチの種類（対比学習、クラスタリングベースの方法、ネガティブフリー方法、アテンションベースの方法）ごとに分類されています。

対比学習アプローチ

対比学習手法は、類似した画像の表現を近づけ、異なる画像の表現を遠ざけることでエンコーダーを訓練します。通常、各画像インスタンスは独自のクラスとして扱われ（インスタンス識別）、バッチ内の他の画像は特徴空間で反発させるためのネガティブ例として機能します。これらの手法は、多くのネガティブ例を提供する大規模なバッチサイズから大きな恩恵を受けます¹。以下は、このカテゴリの主要な論文です：

視覚的表現の対比学習のためのシンプルなフレームワーク（Ting Chen ら, 2020）

著者：Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton
要約：SimCLRという、自己教師付き視覚表現学習のためのシンプルながら効果的なフレームワークを提案します²。この手法は、同じ画像の2つの拡張ビュー（正例ペア）間の合意を最大化しつつ、バッチ内の他の画像を否定例として区別することで画像特徴を学習します。この研究の主要な発見には、強いデータ拡張と学習可能な投影ヘッドの重要性が挙げられます³。特に、SimCLRはバッチサイズ（より多くのネガティブサンプルを提供）とトレーニングステップの増加に伴い性能が向上することを示し、ResNet-50でImageNetにおいて監督学習並みの精度を達成しています¹。バッチ相互作用：SimCLRは、バッチ内の各画像の表現を、同じバッチ内の他のすべての画像の表現（拡張された複製を除く）から遠ざけるバッチ内対比損失（NT-Xent）を使用しています³。つまり、バッチ内の他のすべてのサンプルがネガティブ例として機能し、特定のサンプルの特性学習に直接影響を与えます。このバッチ単位の比較がSimCLRの成功の核心であり、メモリバンクの必要性を排除しています。リンク：[arXiv:2002.05709](https://arxiv.org/abs/2002.05709)（ICML 2020）



教師なし視覚表現学習のための勢いのコントラスト (Kaiming He et al., 2020)

著者: Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, Ross Girshick 要約: 本研究では、モーメンタム更新型エンコーダーを用いて動的な特徴キー辞書を生成する非監督型フレームワーク「MoCo (Momentum Contrast)」を提案する。MoCoは過去のエンコード済み例をキューに保持し、大規模なネガティブサンプルセットとして活用することで、管理可能なバッチサイズでの対比学習を可能にします。辞書キーは、モデルが進化しても一貫性を保つため、移動平均(モーメント)メカニズムで更新されます⁴。MoCoは線形評価においてImageNetで競争力のある結果を達成し、さまざまな検出とセグメンテーションタスクにおいて、非監督型事前学習が監督型事前学習に匹敵または上回る性能を示すことを示しています⁵。バッチ相互作用: MoCoの対比損失は、SimCLRと同様に他の画像の特徴をネガティブとして扱いますが、ネガティブのプールはメモリキューを介して現在のバッチを超えて拡張されます⁴。各反復で、現在のバッチのサンプルはエンコーダーでエンコードされ、モーメンタムエンコーダーでエンコードされたキーの辞書と照合され、InfoNCE損失が計算されます。他のバッチ要素(および最近のバッチからキューに入れられたサンプル)のコンテキストは、モデルが現在の画像を他のさまざまな画像と区別することを促すことで、表現学習に影響を与えます。リンク: [arXiv:1911.05722](https://arxiv.org/abs/1911.05722) (CVPR 2020)

監督付き対比学習 (Prannay Khosla et al., 2020)

著者: Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, Dilip Krishnan 要約: この論文は、対照学習を完全教師あり設定に拡張し、SupCon (Supervised Contrastive) 損失を導入している⁶。SupConは、各インスタンスを独立したクラスとして扱うのではなく、ラベル情報を活用します: 同じクラスラベルを持つ画像の埋め込みは引き寄せられ、異なるクラスラベルを持つものは押し離されます⁷。著者らは、このアプローチが標準的なクロスエントロピー分類損失を上回り、ImageNet精度(ResNet-200でトップ1精度81.4%)とデータ破損に対する頑健性を向上させることを示しています⁸。SupConは、クロスエントロピーと比較して、さまざまなハイパーパラメーターにおいてより安定した性能を示します。バッチ相互作用: SupConは、クラスごとにポジティブサンプルをグループ化することでバッチサンプルの関係を活用します。各バッチ内で、同じクラスのすべての画像は正例クラスターを形成し、他のクラスの画像と対比されます。これにより、各サンプルの表現は、バッチ内に存在する同じクラスのサンプル(引き寄せられる)と異なるクラスのサンプル(押し出される)の両方によって影響を受けます⁷。このバッチ内の複数サンプル対比は、バッチを1対1のインスタンスレベルを超えた正例と負例のミニデータセットとして効果的に活用します。

Link: [arXiv:2004.11362](https://arxiv.org/abs/2004.11362) (NeurIPS 2020)

クラスタリングベースの表現学習

クラスタリングベースの手法は、トレーニング中に類似した画像をグループ化し、画像に疑似ラベルやプロトタイプコードを割り当て、これらのラベルを用いて表現を学習します。これらのアプローチは、バッチコンテキストを活かすため、バッチからオンザフライでクラスターやプロトタイプを形成し、それらのグループ内の一貫性を強制します。以下は、このカテゴリの主要な論文です:

クラスターの割り当ての対比による視覚的特徴の教師なし学習 (Mathilde Caron et al., 2020)

著者: Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, Armand Joulin 要約: オンラインのクラスタリングベースの自己教師学習手法である SwAV (Swapped Assignments) を提案。

9. SwAVは明示的なペアワイズ特徴比較を回避し、代わりにバッチ内の画像をクラスタリングし、同じ画像の異なる拡張におけるクラスタ割り当ての一貫性を強制します 10. 具体的には、各バッチの画像特徴はプロトタイプベクトルのセットにクラスタリングされ、「スワップ予測」タスクが使用されます：モデルは、別の拡張の特徴から1つの拡張のクラスタコードを予測します 11. SwAVは、トレーニング効率を向上させるため、複数の小さなクロップビューを使用するマルチクロップ拡張戦略を導入します 12. この手法は、ResNet-50（線形評価）で75.3%のトップ1 ImageNet精度を達成し、大規模なメモリバンクやモーメンタムエンコーダーを必要とせずに、以前の対照的学習手法と肩を並べます 12 13. バッチ相互作用：SwAVは各反復でバッチの画像特徴を明示的にクラスタリングします 10. バッチ内のクラスタ割り当て（本質的に疑似ラベル）は文脈を提供します：同じクラスタに属する画像は、互いのクラスタコードを予測するように異なる拡張ビューが相互に作用することで、一緒に引き寄せられます。バッチ内の他のサンプルの存在はクラスタプロトタイプに影響を与えます。各サンプルの表現は、バッチ全体の特徴分布から計算されたこれらのプロトタイプ「コード」との関係で学習されます 10. したがって、画像の表現は、他のバッチ画像がそれと共にグループ化されるかどうかに影響を受け、対比学習の利点とクラスタリングを組み合わせます 9.

Link: [arXiv:2006.09882](https://arxiv.org/abs/2006.09882) (NeurIPS 2020)

対比クラスタリング (Yunfan Li ら、2021)

著者：Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey T. Zhou, Xi Peng 要約：本論文は、インスタンスレベルとクラスタレベルの表現学習をオンラインで橋渡しするContrastive Clustering (CC) というアプローチを提案します 14. CCは、インスタンスレベル（データ拡張によるインスタンス識別など）とクラスタレベルの両方で対比的最適化を同時に実行します。鍵となるアイデアは、バッチ特徴量行列の行を「ソフトラベル」（各インスタンスのクラスタへの割り当て確率）とし、列をクラスタプロトタイプとして扱うことです 15 16. インスタンスとクラスタの割り当てレベルで、正例ペアの一致度を最大化し、負例ペアの不一致度を最大化することで、モデルは表現とクラスタ割り当てを同時に学習します 17. CCのエンドツーエンドトレーニングは、従来の深層クラスタリング手法に比べて大幅に改善されたクラスタリング性能（例：CIFAR-10/100でのNMIが大幅に上昇）を実現します 18. バッチ相互作用：対比クラスタリングは、バッチ単位でのクラスタ割り当ての計算に依存しています。各バッチにおいて、特徴量を用いてすべてのサンプルの「ソフト」クラスタ割り当てを計算し、クラスタレベルの対比損失が特徴量をバッチのクラスタ中心点に引き寄せつつ、他のクラスタから押し出します 14 19. 同時に、通常のインスタンスレベルの対比損失は、バッチ内の他の画像を否定例として考慮します。クラスタ割り当て自体は、バッチの特徴量分布（通常はk-meansやバッチ上の割り当てアルゴリズム経由）から導出されるため、各サンプルの学習信号は他のバッチサンプルのグループ化方法に直接影響を受ける。リンク：AAAI 2021 論文 (AAAI 2021)

（注：このアプローチのもう1つの注目すべき方法は、Liら (ICLR 2021) のPrototypical Contrastive Learning (PCL) で、各画像に複数のプロトタイプ中心を割り当て、画像とその最も近いプロトタイプを対比させることで、対比学習とクラスタリングを組み合わせたものです。これは、バッチレベルクラスタリングやプロトタイプを特徴学習のガイドとして使用する傾向をさらに強調しています。)

ネガティブフリー・シアーメーズ表現学習

対比学習手法は通常、バッチ内の他のサンプル（ネガティブ）との比較に依存しますが、最近の「ネガティブフリー」シアーメーズネットワークは、明示的なネガティブペアなしで意味のある表現を学習できることを示しています。これらの手法は依然として複数のサンプル（通常は同じ画像の2つの拡張ビュー）を活用し、バッチ統計やアーキテクチャのトリックを組み込むことで単純な解を回避しています。影響力のある論文をいくつか紹介します：

Bootstrap Your Own Latent : A New Approach to Self-Supervised Learning (JeanBastien Grill et al., 2020)

Authors: Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, et al. (13 その他の手法) 要約: 本論文は、驚くべきことにネガティブサンプルや対比ペアを使用しない先駆的な自己教師付き学習手法BYOLを提案します²²。BYOLは2つのネットワークを採用します: オンラインネットワークとターゲットネットワーク。同じ画像の2つの拡張ビューが与えられると、オンラインネットワークはターゲットネットワークがもう一方のビューの表現を予測するように訓練されます²³。ターゲットネットワークのパラメーターは、オンラインネットワークのパラメーターの遅延移動平均(モーメンタムエンコーダーの一種)です。このブートストラップメカニズムにより、BYOLは崩壊を回避し、高品質な特徴を学習し、明示的なネガティブサンプルなしでResNet-50で74.3%のトップ1 ImageNet精度(より大きなResNetでは79.6%)を達成します²⁴。転移学習と半教師付きベンチマークで最先端性能に匹敵または上回っており、対比的ネガティブが表現学習に厳密に必要なことを示しています。バッチ相互作用: BYOLのトレーニング目標は同じ画像からの拡張ペアに焦点を当てているため、バッチ内の他の画像と直接的に引き合いや押し合いは起こりません。ただし、バッチ正規化や他の暗黙の効果により、バッチは依然として文脈を提供します(例: BN層はバッチ全体で統計を計算します)。重要な点は、BYOLが他のバッチサンプルと比較せずに特徴表現を学習できるケースを示していることです - 他の画像の存在は正規化のためだけに偶発的です。オンラインネットワークにおけるターゲットネットワークのモーメント更新と予測ヘッドへの依存が崩壊を防止しており、ネガティブサンプルの比較ではありません²⁴。この方法は、バッチ単位の対比の代替案として、2つのネットワークと反復的な知識蒸留を them.

Link: [arXiv:2006.07733](https://arxiv.org/abs/2006.07733) (NeurIPS 2020)

Barlow Twins : Self-Supervised Learning via Redundancy Reduction (Jure Zbontar et al., 2021)

著者: Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, Stéphane Deny 要約: Barlow Twinsは、負例なしで不変特徴を学習するための冗長性削減に基づく新しい損失関数を提案します²⁵。この方法は、各画像の2つの拡張ビュー(2つの同一ネットワークで処理されたもの)を取り、バッチ内の2つの埋め込みセットのクロス相関行列を計算します。目的関数は、このクロス相関を単位行列にできるだけ近づけるように強制します²⁶。直感的には、これは次のことを意味します: (1) 対角成分(同じ画像、同じ特徴成分)は1に近づけられ、同じ画像の2つのビュー間の類似性が促進され、(2) 非対角成分は0に近づけられ、異なる特徴次元が相関を解消し、冗長性が削減されます²⁷。これにより、ネットワークがすべての特徴を同じようにエンコードする単純な定数解を回避します。Barlow Twinsは、ImageNetの線形評価において当時の最良手法と同等の性能を達成し、大規模バッチや特別なモーメンタムエンコーダー、ストップグラディエントを必要とせずに実現しました²⁸。バッチ相互作用: Barlow Twinsの核心は、バッチ全体で計算されるクロス相関行列です - この行列はバッチ内のすべてのサンプルの表現間の関係を捕捉します。損失関数はバッチ統計に明示的に依存しています: バッチ内の2つの異なる画像が相関する特徴次元を持つ場合、対角線外の要素がその関係を捕捉し、損失が罰則を課します。したがって、明示的なネガティブペアがなくても、バッチ内の他のサンプルの存在が対角線外の項を通じて損失に影響を与えます。各画像の特徴は、その拡張双子と一致するように最適化されるだけでなく、各次元において他の画像の特徴と相関しないように最適化されます(バッチ全体での期待値において)²⁵。要約すると、Barlow Twinsはバッチレベルの共分散情報を用いて表現を豊かにします。リンク: [arXiv:2103.03230](https://arxiv.org/abs/2103.03230) (ICML 2021)

VICReg：自己教師付き学習のための分散-不変性-共分散正規化 (Adrien Bardes et al., 2022)

Authors: Adrien Bardes, Jean Ponce, Yann LeCun

要約：VICRegは、明示的な正則化項を用いて表現の崩壊を回避する別のネガティブフリー自己教師付き学習法です²⁹。3つの項を導入します：(1) 不変性項(2つの拡張の埋め込み間の平均二乗誤差)、埋め込み間の平均二乗誤差、BYOLと同様の類似性を促進)、(2) バッチ内の埋め込みの次元ごとの分散が閾値を超えるように強制する分散項(すべての出力が定数に崩壊するのを防止)²⁹、および(3) Barlow Twins に類似し、異なる特徴次元間の非対角共分散(相関除去)を罰する共分散項³⁰。これらの要素を組み合わせることで、VICRegはImageNetを含むタスクで最先端の性能と安定した学習を実現する。著者らは、他の手法(BYOLO/SimCLRなど)に分散正規化を追加することで、その安定性と性能を向上させられることを示している³¹。(VICRegはICLR 2022で採択された。)

バッチ相互作用：VICRegは損失関数にバッチ統計量を明示的に使用しています。分散項はバッチ内の埋め込みを分析し、各特徴次元においてバッチに十分な分散があることを確認します(つまり、画像がすべて同じ値を生成していないこと)²⁹。共分散項は、埋め込みのバッチの共分散行列(中心化後)を計算し、対角線以外の要素を罰則化します。つまり、バッチ内のサンプル間で2つの異なる次元が相関した応答を示す場合、それらをゼロに近づけます²⁹。これらのメカニズムは情報的多様性を確保し、バッチ内の関係性を暗黙的に利用します(分散や共分散を測定するには複数のサンプルが必要だからです)。したがって、VICRegは異なる画像インスタンスを対比して比較するのではなく、バッチレベルのダイナミクス(分散と共分散)を利用して特徴空間を形作ります。リンク：arXiv:2105.04906 (ICLR 2022)

注：もう1つの重要なネガティブ不要手法はSimSiam (Xinlei Chen & Kaiming He, 2021)で、単純なSiameseネットワークがネガティブデータ、大規模バッチ、モーメンタムエンコーダーなしで良い特徴を学習できることを示しました。これは2つのビューとストップグラディエントトリックのみを使用しています³²。SimSiamの成功は、バッチコンテキストを最小化できる場合があることをさらに確認していますが、上記のほとんどの手法はバッチ相互作用をより明示的に活用しています。

注意ベースのバッチ相互作用

損失関数とクラスタリングを超えて、研究者はバッチ内のサンプル間で注意を直接許可するアーキテクチャモジュールを、特徴抽出を強化するために探索してきました。これらの方法は、バッチ全体を観察することで重要なサンプルや特徴を特定することを目的としています。以下は注目すべき例です：

BA²M：バッチ意識型アテンションモジュールによる画像分類 (Qishang Cheng et al., 2021)

著者：Qishang Cheng, Hongliang Li, Qingbo Wu, King Ngi Ngan 要約：BA²Mは、個々の画像内ではなく画像のバッチ全体で動作する新しい注意メカニズムを導入します³³。従来の注意モジュール(SEブロック、CBAMなど)は、1つのサンプル内の特徴強化に焦点を当てています。これに対し、BA²Mはバッチ意識のステージを追加します：まず、各画像内でマルチスケール注意マップ(チャンネル別、局所空間、グローバル空間)を融合して、各画像のSample-wise Attention Representation (SAR)を計算します³⁴。次に、バッチ内のすべてのSARを正規化関数に投入し、各サンプルの重みを算出します³⁵。これらの重みは、バッチの文脈における各サンプルの特長の相対的な重要性または複雑さを示します³⁶。ネットワークはこれらの重みを使用して特徴量を再加重または調整し、バッチ内のより困難なまたは情報量の多い例に注意を向けることができます。

CIFAR-100とImageNet-1Kでの実験では、さまざまなCNNアーキテクチャにおいて一貫した性能向上が確認され、古典的なアテンションモジュールやサンプル再加重手法を凌駕しました 37。バッチ相互作用：BA²Mは、バッチ内の異なるサンプルが互いに注意を向けるように明示的に設計されています。バッチ内の各サンプルの注意特徴量をバッチレベルで正規化することで、トレーニングバッチ内のサンプル間の特徴の重要性を区別します 35。例えば、バッチ内の1つの画像がより複雑または珍しい特徴を含む場合、BA²Mはその画像に高い重みを割り当て、そのフォワードパス中にネットワークがその画像に焦点を当てるようにします。これは、与えられた画像の表現が孤立して調整されるのではなく、同じバッチ内の他の画像の文脈において調整されることを意味します。本質的に、モジュールはバッチ全体の特徴分布に対して特徴を強調または抑制する方法を学習し、バッチ内のサンプル間の注意ベースの相互作用を導入します 33。リンク：arXiv:2103.15099（2021年プレプリント）

Conclusion

要約すると、画像の表現学習における最近の進展は、より豊かな特徴表現を抽出するためにバッチレベルの関係を increasingly 活用しています。対比学習手法は、他のバッチサンプルを明示的なネガティブ（または監督学習の場合のポジティブ）として使用し、埋め込み空間を形作る。クラスタリングベースのアプローチは、各バッチ内でオンザフライのグループを形成し、インスタンスのみのビューよりも広い文脈を提供する。新しいネガティブフリー技術は、インスタンスの直接比較を回避しつつ、バッチ統計（共分散など）やシアーメイズアーキテクチャを活用して不変性を学習する。最後に、BA²Mのようなアテンションメカニズムは、ネットワークのフォワードパスにバッチコンテキストを直接組み込むことで重要なサンプルに焦点を当てる可能性を示しています。これらの研究は、トレーニング中に複数のサンプルのコンテキストを同時に考慮することで、より識別能が高く頑健な画像表現が得られ、その結果、画像分類ベンチマークでの性能が向上することを示しています。この分野は引き続き進化しており、トレーニングサンプル間の関係をさらに活用する新たな手法の探索が進んでいます。

1 2 3 [2002.05709] 視覚的表現の対比学習のためのシンプルなフレームワーク h
https://arxiv.org/abs/2002.05709

4 5 [1911.05722] Momentum Contrast for Unsupervised Visual Representation Learning
https://arxiv.org/abs/1911.05722

6 7 8 [2004.11362] Supervised Contrastive Learning
https://arxiv.org/abs/2004.11362

9 10 11 12 13 papers.neurips.cc
https://papers.neurips.cc/paper_files/paper/2020/file/70feb62b69f16e0238f741fab228fec2-Paper.pdf

14 15 16 17 18 19 対
比クラスタリング Contrastive Clustering
https://yunfan-li.github.io/assets/pdf/Contrastive%20Clustering.pdf

20 21 ope
nreview.n
et openreview.net
https://openreview.net/references/pdf?id=8Gyq2JakVU

22 23 24 [2006.07733] Bootstrap your own latent: A new approach to self-supervised Learning
https://arxiv.org/abs/2006.07733

25 26 27 28 [2103.03230] Barlow Twins: Self-Supervised Learning via Redundancy Reduction
https://arxiv.org/abs/2103.03230

29 30 31 [2105.04906] VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning
<https://arxiv.org/abs/2105.04906>

32 [2011.10566] Exploring Simple Siamese Representation Learning
<https://arxiv.org/abs/2011.10566>

33 34 35 36 37 [2103.15099] BA²M: A Batch Aware Attention Module for Image Classification
<https://arxiv.org/abs/2103.15099>