

バッチ間の関係性を活用した特徴表現強化の最新研究

近年、バッチ内の他サンプル情報を活用して単一画像の特徴表現を強化する手法が注目されています。従来の注意機構は各サンプル内部の特徴強調に留まり、サンプル間の差異を無視しがちでした¹。これに対し、BA²M (Chengら 2021年) はバッチ全体を考慮した注意機構を導入し、異なる画像間の関係性を特徴抽出に組み込むことで性能向上を達成しました¹。2020年以降、このBA²Mの理念を発展させ、バッチ内相互作用を活用した新たなアーキテクチャや損失関数が数多く提案されています。以下、主な論文についてタイトル、著者・年度、概要、バッチ関係性の扱い、および論文リンクをまとめます。

BA²M: A Batch Aware Attention Module for Image Classification (Chengら, 2021年)

- **概要:** 従来の注意層が各画像内部の特徴にのみ集中しサンプル間の判別性を考慮しない問題に着目し、BA²Mはバッチ対応型注意モジュールを提案しました¹。各画像についてチャンネル・ローカル空間・グローバル空間の注意マップを融合した特徴 (SAR) を計算し、バッチ内全サンプルのSARを正規化関数に入力してサンプルごとの重みを算出します²。この重みにより、同一バッチ内でコンテンツの複雑さが異なる画像間で特徴の重要度を調整します³。モジュールは軽量で様々なCNNに組み込み可能で、ネットワークとエンドツーエンドで学習できます⁴。
- **バッチとの関係性:** バッチ内の全画像の特徴表現をまとめて評価し、それぞれのサンプルに重み付けする点が特徴です⁵。これにより内容が複雑な画像は重みを上げ、容易な画像は下げるように学習し、サンプル間の違いを特徴抽出に反映させます³。従来の「難易度に基づくサンプル再重み付け手法」を上回る効果を示しました⁶。
- **評価:** CIFAR-100やImageNet-1Kで検証した結果、BA²Mの導入により様々なネットワークの画像認識精度が向上し、既存の注意機構より高い性能を示しました⁶。特に古典的な注意手法や損失値に基づくサンプル重み付けを明確に上回ることが報告されています⁶。
- **論文リンク:** arXiv:2103.15099⁶

Batch Transformer: Look for Attention in Batch (Herら, 2024年)

- **概要:** 顔表情認識 (FER) の性能劣化要因であるノイズや曖昧さに対処するため提案された手法です。Batch Transformer Network (BTN) は、ViTベースのモデルにクラス・バッチ注意 (CBA) モジュールとマルチレベル注意 (MLA) を組み込み、単一画像ではなくバッチ内の複数画像の特徴に基づいて信頼性の高い情報を抽出します⁷。CBAは各画像のクラス予測情報を他の画像に反映させ、MLAは特徴レベル間の相関を捉えることで過学習の抑制にも寄与します⁷。実験では複数のFERベンチマークで既存手法を安定して上回る高精度を達成しました⁷。
- **バッチとの関係性:** CBAモジュールによりバッチ内のデータ同士が互いに注意を及ぼし合う仕組みを導入しています⁸。具体的には、特徴の類似した複数画像のクラス予測結果を互いに参照して特徴を調整し、同一クラス内では特徴を近づけ、異なるクラス間では特徴をより分離するよう作用します⁹。これはBA²M同様、「バッチ内で情報を共有して各画像の特徴信頼性を高める」発想に基づいています。
- **評価:** 提案手法を用いたBTNはRAF-DBやFERPlusなどFERデータセットにおいて最新の既存手法を上回る最高精度を記録しました⁷¹⁰。例えば従来SOTA手法との比較で平均精度を僅かではありますが着実に向上させ、FERタスクにおける有効性が示されています¹¹。
- **論文リンク:** arXiv:2407.04218⁷

IBAFFormer: Intra-Batch Attention Transformer for Domain Generalized Semantic Segmentation (Sunら, 2023年)

- **概要:** ドメイン一般化セマンティックセグメンテーション(DGSS)において、ソースドメインのみで学習したモデルが未知のターゲットドメインにも適応できるようにする研究です。IBAFFormerではTransformerの注意機構を拡張し、各注意ブロックでバッチ内の他画像からの情報を取り込むことで、特徴表現の多様化と文脈強化を図りました¹²。具体的には平均ベース (MIBA) と要素ベース (EIBA) の2種類のインラバッチ注意を提案し、自己注意に加えて異なる画像間の相関関係を捉えるようにしています¹³。これらを統合したIBAFFormerはドメインシフト下でも優れた汎化性能を発揮しました¹⁴。
- **バッチとの関係性:** 各バッチ内の独立したサンプル間の相関をTransformer内部で活用する点が特徴です¹²。従来のTransformer型セグメンテーションは各画像内の自己注意に留まりましたが、本手法では同じバッチ内の他画像の特徴をキー/バリューとして参照する注意機構を導入しています¹²。このバッチ内注意によって、BA²Mが達成したようなサンプル間関係に基づく特徴強調をセグメンテーションタスクにも適用し、より頑健な特徴抽出を可能にしています。
- **評価:** 提案手法を実装したIBAFFormerは、ドメイン一般化セグメンテーションの最先端性能を達成しました¹⁵。多様なドメイン間で学習・評価する実験で既存手法を上回るmIOUを示し、各構成要素も消去法検証により有効性が確認されています¹⁶。
- **論文リンク:** arXiv:2309.06282¹⁴

Supervised Contrastive Learning (Khoslaら, 2020年)

- **概要:** 表現学習におけるコントラスト学習手法を教師あり設定に拡張した研究です。従来のコントラスト学習 (例: SimCLR) は自己教師でバッチ内の他画像をネガティブ例として活用していましたが、本手法ではクラスラベルを活用し、同一クラスの画像特徴同士を引き寄せ、異なるクラス同士を遠ざける損失関数 (SupCon損失) を提案しました¹⁷。これにより同じクラスの画像は埋め込み空間でクラスターを形成し、クラス間の境界がより明確になるよう学習されます¹⁷。ResNet-200を用いたImageNet実験では従来の交差エントロピーより高い81.4%のTop-1精度を達成し、他のデータセットやモデルでも安定した精度向上を示しました¹⁸。また、自然画像汚損に対する頑健性向上やハイパーパラメータ感度の低減といった利点も報告されています¹⁹。
- **バッチとの関係性:** バッチ内のサンプル間の関係を積極的に損失設計に組み込んだ点が特徴です。具体的には、各アンカー画像に対し同じクラス他バッチ例すべてをポジティブ対例、異なるクラスの例をネガティブ対例として扱うことで、バッチ内の全関係を考慮した学習を行います¹⁷。これはBA²Mのようなアーキテクチャ上の工夫ではなく損失関数側のアプローチですが、「バッチ内でのクラスターリングを促進して判別的特徴空間を形成する」という点で目的を同じくしています。
- **論文リンク:** arXiv:2004.11362²⁰

SwAV: Unsupervised Learning of Visual Features by Contrasting Cluster Assignments (Caronら, 2020年)

- **概要:** 教師なし対比学習にクラスターリングの考え方を取り入れた手法です。通常のコントラスト学習はバッチ内のペア比較を多数行うため計算コストが高いですが、SwAVではオンラインクラスターリングによってバッチ内のデータをグループ化し、異なるビュー (データ拡張) 同士のクラスター割当てが一致するよう予測させることで特徴表現を学習します²¹。具体的には、一方の画像ビューから得た表現で他方のビューのクラスター所属を予測する (スワップ予測) 損失を導入し、直接のペア比較を回避しました²²。この手法は大規模メモリバンクや極端に大きなバッチサイズを必要とせず、小規模バッチでも効率良く学習可能です²³。さらにマルチクロップと呼ばれる多数の小さなクロップ画像を含む拡張も組み合わせ、追加の計算負荷なく多様なビューから学習できる工夫も導入されています

24。結果として、ImageNetデータでResNet-50のトップ1精度75.3%を達成し、教師あり学習を含む従来法に匹敵・一部上回る性能を示しました 25。

- **バッチとの関係性:** クラスタリングはバッチ内の全サンプルの相対関係に基づいて行われるため、SwAVではバッチ全体の特徴分布を活用して各画像の表現を更新しています 21。Sinkhorn-Knoppアルゴリズムにより各バッチで画像からプロトタイプへの割り当てを計算し直す際、**バッチ内の画像同士がプロトタイプ数に均等に割り振られるよう制約**することで、多様なサンプル間関係を反映した学習が可能です（他サンプルが存在しないとクラスタ割当て自体が定義できません）。このように**暗黙的にバッチ全体の情報を用いて特徴空間を整理する**アプローチであり、BA²Mとは手法の種別（教師なし vs 注意機構）は異なりますが「他データからの情報で単一画像表現を良くする」という点で通底しています。
- **論文リンク:** arXiv:2006.09882 26

Barlow Twins: Self-Supervised Learning via Redundancy Reduction (Zbontarら, 2021年)

- **概要:** 自己教師あり学習で**表現の冗長性を削減**することにより高性能化を図った手法です。対比学習では**ネットワークが定数出力に崩壊するのを避ける工夫**が必要ですが、Barlow Twinsでは**同一画像の2種類の拡張入力から得た出力のクロス相関行列を計算し、それを単位行列に近づける**という新しい目的関数を導入しました 27。対角成分を1（同一画像ペアは完全相関）、非対角成分を0（異なる特徴次元同士は無相関）にすることで、**同じ画像から得た特徴ベクトル同士は極力類似させつつ、特徴次元間の情報重複を排除**します 28。この手法は**巨大バッチや特殊なアーキテクチャ拡張（予測器ネットワークやモーメント更新など）を不要**とし、高次元の特徴ベクトルほど効果を発揮する点も特徴です 29。ImageNetでの線形分類評価では当時の最先端手法に匹敵する精度を示し、ラベル付きデータが少ない状況での半教師あり学習でも既存法を上回る性能を報告しています 30。
- **バッチとの関係性:** 提案手法のクロス相関行列は**バッチ内の全サンプル対について計算**されます。具体的には、あるバッチに含まれる画像の2ビューから得た特徴 $\{Z_A\}$ と $\{Z_B\}$ について、 $\{Z_A\}^T Z_B$ のようにバッチ内平均で相関を求めるため、**各サンプルの寄与が全体として統計的に評価**される仕組みです 27。このように**バッチ全体をまたいだ統計（共分散）**に基づき特徴表現を正則化する点で、「バッチ内他サンプルとの関係性を損失に反映させる」一例と言えます。BA²Mが注意モジュール内でサンプル間の重み付けを行うのに対し、Barlow Twinsでは**損失関数上でサンプル間相関を制御**しており、アプローチは異なりますが目的は共通しています。
- **論文リンク:** arXiv:2103.03230 28

CC-Loss: Channel Correlation Loss for Image Classification (Songら, 2020年)

- **概要:** 特徴量分布とモデル構造（チャンネル）の関係性に着目して設計された**新規損失関数**です。従来の分類ロス（交差エントロピー等）に**クラス内・クラス間の特徴分離性を高める工夫**を加えた手法はいくつかありましたが、本手法ではさらに一歩進めて「**クラス**」と「**チャンネル**」間の**対応関係を明示的に活用**しています 31。具体的には、各訓練サンプルに対し**チャンネル注意ベクトル**（そのサンプルの特徴マップに対するチャンネル重要度）を計算し、**同じクラスのサンプルではこのチャンネル注意ベクトルが一致し、異なるクラス間では距離が離れる**ようユークリッド距離に基づく損失を定義します 32。結果として、**クラス内では共通するチャンネルが強調されクラス間では異なるチャンネルが活性化**するよう学習が進み、チャンネル次元で見た特徴の類似度にクラスごとの特徴的パターンが現れます 32。
- **バッチとの関係性:** 学習時には**バッチ内の複数サンプル間で上記のチャンネル相関制約をかける**ことで、各クラスの特徴チャンネル分布をそろえるようにしています 32。例えば、バッチ中の同一クラス画像について計算されたチャンネル注意ベクトル同士の距離を小さく、異なるクラス間では大きくするよう損失を加えることで、**ミニバッチ内でクラス毎に特徴の軸（チャンネル）を揃える効果**があります

³²。このようにバッチ内のサンプルを対比させて特徴表現を正則化する発想はBA²Mとも通じるものですが、実装形態は異なりこちらは損失関数として組み込まれます。

- **評価:** 提案されたCC-Lossを組み込むことで、ResNetなど2種類のバックボーンを用いた画像分類モデルがCIFAR-100やImageNet-LT等のデータセットで**既存の最先端損失関数を上回る精度**を達成しました³³。クラス不均衡や微細分類といった難しい設定においても、チャンネル相関に基づく特徴強化が分類性能向上に有効であることが示されています。

- **論文リンク:** arXiv:2010.05469 ³⁴

The Bad Batches: Enhancing Self-Supervised Learning in Image Classification Through Representative Batch Curation (Göksu & Pugeault, 2024年)

- **概要:** 自己教師あり対比学習において、ランダムなデータ拡張に依存するがゆえに**不適切なペア (false positive/negative)** が生じ学習を不安定化させる問題に対処した研究です。著者らはこれら「悪いバッチ」による影響を緩和するため、**フレシェ距離**に基づくペア類似度計算でバッチ内の画像ペアを評価し、**代表的な正例ペアのみで構成されたバッチをキュレーション**する手法を提案しました³⁵。具体的には、事前学習したResNetを用いて画像間の距離 (**Fréchet ResNet Distance**) を測定し、距離の近い画像ペアを高信頼のポジティブ対とみなしてバッチに組み入れます³⁵。これによりランダム拡張に頼った場合に比べ誤ったポジティブ・ネガティブ対が減少し、対比学習の収束が速まり小さなバッチサイズでも高性能な表現が得られるようになります³⁶³⁷。
- **バッチとの関係性:** 本手法は直接的にモデルの構造を変えるものではありませんが、**バッチを構成するサンプル選択に他サンプルとの距離情報を利用する**点で「バッチ内関係性の活用」に該当します。すなわち、バッチ内のペア類似度に応じて学習に**有益なサンプルを選別・重み付け**することで、各画像の表現学習を間接的に高めています³⁸。このアプローチはBA²Mのようなネットワーク内注意とは異なりますが、「他の画像との関係を見ることで単一画像の表現の質を上げる」という目的は共通しています。
- **評価:** 提案手法で自己教師あり学習したモデルを線形分類器で評価したところ、STL-10で**トップ1精度 87.74%**、Flower102で**99.31%**と非常に高い精度を達成しました³⁹。従来法に比べ小規模データセットでも高精度を発揮し、自己教師あり画像分類の最先端を押し広げる可能性が示されています³⁹。なおこの研究はIJCNN 2024で発表されており、今後さらなる発展が期待されます。
- **論文リンク:** arXiv:2403.19579 ⁴⁰

<small>各論文の引用箇所は該当するarXiv概要や論文中の記述を示しています。【】内の番号は出典を示します。</small>

1 2 3 4 5 6 [2103.15099] BA²M: A Batch Aware Attention Module for Image Classification

<https://arxiv.org/abs/2103.15099>

7 8 9 10 11 Batch Transformer: Look for Attention in Batch

<https://arxiv.org/html/2407.04218v1>

12 13 14 15 16 [2309.06282] IBAFormer: Intra-batch Attention Transformer for Domain Generalized Semantic Segmentation

<https://arxiv.org/abs/2309.06282>

17 18 19 20 [2004.11362] Supervised Contrastive Learning

<https://arxiv.org/abs/2004.11362>

21 22 23 24 25 26 [2006.09882] Unsupervised Learning of Visual Features by Contrasting Cluster Assignments

<https://arxiv.org/abs/2006.09882>

27 28 29 30 [2103.03230] Barlow Twins: Self-Supervised Learning via Redundancy Reduction

<https://arxiv.org/abs/2103.03230>

31 32 33 34 [2010.05469] CC-Loss: Channel Correlation Loss For Image Classification

<https://arxiv.org/abs/2010.05469>

35 36 37 38 39 40 The Bad Batches: Enhancing Self-Supervised Learning in Image Classification Through Representative Batch Curation

<https://arxiv.org/html/2403.19579v1>