

空間ラベリング：非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

張家銘* 李佳憲† 五十嵐健夫*

概要. 非専門家による画像注釈において、ラベルの品質は重要な問題である。本研究では、非専門家の画像注釈のラベル品質を改善するために空間レイアウトを活用することを提案する。本システムでは、アノテーターは最初にターゲット画像を空間的にレイアウトし、関連するアイテムと一緒に配置しながらオープンスペースでラベル付けを行う。これは、暫定的なラベリングのための作業スペース（空間組織）として機能する。作業中に、アノテーターはアイテム間の類似点と相違点を観察して整理する。最後に、アノテーターは空間レイアウトの結果に基づいて画像に最終的なラベルを割り当てる。画像ラベリングタスクを用いて本システムと従来の非空間レイアウトを比較するユーザー調査を行った。結果として、アノテーターは非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを使用することで、より正確にラベリングタスクを完了できることを示した。

1 はじめに

非専門家による画像のラベル付けは非常に困難であり、多くのエラーが含まれている可能性がある。したがって、プロセスをより効率的にし、エラーを減らすためのサポートを提供することが重要である。例えば、犬の品種に関する十分な知識のないアノテーターが、犬の品種のラベルを犬の画像に割り当てるように求められた場合、ラベルに関連付けられたサンプル画像を参照することができるが、サンプル画像では類似した品種間の微妙な違いを捉えるには不十分な場合がある。このような場合、アノテーターは、サンプル画像とラベルを割り当てる画像を比較することにより、微妙な違いを観察および整理できる。この観察と整理のプロセスは、非専門家によるラベリングにおいて重要な役割を果たしていると考えられるが、従来の画像ラベリングツールでは十分にサポートされていない。本研究では、アノテーターの空間レイアウトを活用して、画像とラベルの類似点と相違点の観察および整理することをサポートする。この空間編成プロセスは、暫定的なラベリングとして機能する。

図1は、プロトタイプの実装のスクリーンショットである。概念的な類似性を表す画像とラベルを空間的にレイアウトするためのオープンスペースがアノテーターに提示される。アノテーターはターゲット画像をオープンスペースにドラッグし、可能なラベルを近くに配置する（つまり、画像はラベルの1

つに属している可能性がある）。このプロセスにより、アノテーターは特定のドメインで画像のラベルをより正確に選択できるようになると考えられる。



図1. 空間ラベリングプロトタイプ実装のスクリーンショット。

本研究の空間レイアウトラベリングと従来の非空間レイアウトラベリングを比較するため、画像ラベリングタスクを用いてユーザー実験を行った。結果として、参加者は非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを介して特定のラベリングタスクをより正確に完了したことが示された。さらに、参加者は、非空間レイアウトラベリングタスクよりも空間レイアウトラベリングタスクで選択したラベルにより自信を持っていた。これらの調査結果は、空間レイアウトインターフェイスが、ラベル付けプロセス中にオープンスペースに画像とラベルをレイアウトすることにより、専門家以外の画像注釈のラベル品質を向上できることを示している。

本論文は、ACM CHI 2021 で発表された論文の内容をまとめたものである[17]。

2 関連研究

手動の画像注釈（例：VIA [1]、LabelMe [2]、ESP [3]および階層的タスク割り当て[4]）を支援し、非専門家の注釈（例：Revolt [5]）のラベル品質を向上させるために、さまざまな注釈ツールが提案されている（Pairwise HITS [6]、2ラウンドアノテーションフレームワーク [7]）。これらのツールのほとんどは、労働負荷の高いアノテーション作業のプロセスとラベルの品質を向上のために、より支援的で効率的かつ楽しむことができるシステムを提供している。しかし、これらのツールは通常、アノテーターが十分なドメイン知識を持っていることを前提としている。さらに、十分なドメイン知識が不足している専門家以外のアノテーターのサポートはまだ制限されており、改善する必要がある。クレシャら[8]は、Webページ分類のためのデータラベリングにおける概念進化という考えを導入し、データラベリング中にアノテーターが概念を定義および改良するのに役立つ2つの構造ラベリングソリューションを提示した。本研究とは、データラベリングにおけるアノテーターのコンセプト編成をサポートするという同じ動機を共有している。ただし、クレシャら[8]は主にWebページ分類のコンテキストでの時間の経過に伴う概念の変化に焦点を当てている一方、本研究は画像分類のコンテキストでの単一セッションでのアイテムの観察と概念の整理に焦点を当てている。

空間レイアウトの概念は、情報ナビゲーション用のズーム可能なインターフェイスなど、さまざまな研究で使用されてきた[9、10]。バウアーら [11]は、ズーム可能なインターフェイスの概念と「パイル」メタファーに基づいて個人情報収集を管理するための空間ツールを提案した。渡辺ら[12]は、「バブルクラスター」という名前のインターフェイスを導入したが、これはオブジェクトのグループ化とグループ化解除を通じて、グラフィックオブジェクトの空間的集約を操作するための手法である。人々が空間レイアウトインターフェイスを介して情報やドキュメントをより効果的に管理できる理由の1つとして、空間記憶（つまり、環境内の何かを記憶する能力）が挙げられる[13]。空間レイアウトの概念は、アノテーターが画像とラベルの類似点と相違点を観察および整理するためのラベリングインターフェイスデザインにも使用できると考えられる。さらに、空間レイアウトは、キャンバス上でのユーザーインタラクションを活用して検索プロセスを容易にする検索システムでも使用されている[14、15]。対照的に、本研究では、ユーザーが手動で画像に注釈を付ける際に考えを整理するためのワークスペースとして、空間レイアウトインターフェイスを使用する。空間レ

アウトは、システムによって作成または解釈はされない。

3 提案手法

本研究のラベリングインターフェイスは、空間レイアウトの概念に基づいて設計された。本システムは、ラベリングプロセス中に概念的な類似性を表す画像とラベルを空間的にレイアウトするためのオープンスペースを提供する。図2に、インターフェイスの初期状態を示す。ラベル（各ラベルにはテキストで表示される名前とサンプル画像が含まれている）がインターフェイスの両側に一覧表示され、ラベル付けされるターゲット画像が中央に1つずつ表示される。

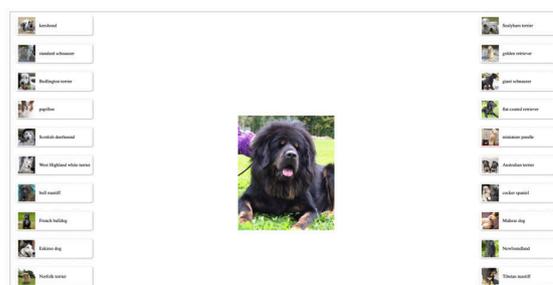


図2. 空間ラベリングインターフェイスの初期状態。

図3は、インターフェイスの動作状態を示している。アノテーターは、画像のラベルを選択する前に、オープンスペースに画像とラベルを空間的に配置する（たとえば、関連するアイテムを一緒に配置する）。



図3. 空間ラベリングインターフェイスの動作状態. 空間レイアウトインターフェイスには、アノテーターがラベル付けタスクを完了するための4つの主要な機能が含まれている：(a) 画像とラベルを空間的にレイアウトする、(b) 画像にラベルを割り当てる、(c) 確信度を示す、(d) 割り当てられたラベルを変更する（図4を参照）。

(a) 画像とラベルを空間的にレイアウトする
オープンスペースでは、アノテーターは概念的な類似性を表すために画像とラベルをレイアウトする。

空間ラベリング：非専門家による画像注釈のラベル品質を改善するための空間レイアウトの活用

より類似した画像がグループ化され、候補となるラベルが近くに表示される。

(b) 画像へのラベルの割り当て

オープンスペースに画像とラベルを配置した後、アノテーターは空間レイアウトの結果に基づいて画像に最終的なラベルを割り当てる。

(c) 信頼状態を示す

画像にラベルを割り当てた直後（ドラッグアンドドロップ操作）、ポップアップウィンドウが表示され、アノテーターにラベル選択の確信度を示すように求められる。

(d) 割り当てられたラベルの変更

ラベルをクリックするとそのラベルが割り当てられたすべての注釈付き画像を表示される。アノテーターは、画像を別のラベルまたはオープンスペースにドラッグアンドドロップすることで、既に割り当てられているラベルを変更できる。

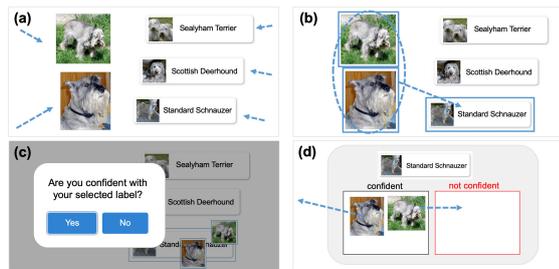


図 4. 空間レイアウトラベリングインターフェイスのユーザーインタラクション

4 ユーザー実験

16人の参加者（18～49歳の男性8人と女性8人）が、ユーザー実験に参加した。参加者は全員、画像ラベリングまたは画像分類に関する専門知識の経験はなかった。また、全参加者がペットとして犬を飼っておらず、犬のカテゴリーについての専門知識を持っていなかった。

ラベル付けタスクには、ImageNet (ILSVRC 2012) [16]の画像データセットを使用した。まず、犬のラベル（品種）ごとに100枚の画像をランダムに選択し、合計12,000枚の画像を基本データセットとして使用した。次に、同等の難易度の2つの互いに素なデータセット（データセットAとデータセットB）を手動で作成した。各データセットには、20のラベル（20の犬種）と50の画像（選択した20の犬種に属する50の犬の画像）が含まれていた。2つのデータセットのラベルと画像の間に重複はなかった。図5は、3つの難易度レベルでデータセットAとデータセットB用に選択された20の犬種（ラベル）を示している。

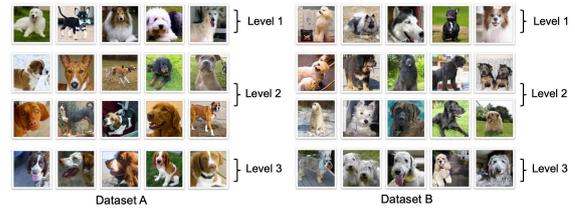


図 5. データセット A およびデータセット B で使用される 20 の犬種（ラベル）。レベル 1：明確。レベル 2：あいまい。レベル 3：非常にあいまい。

ユーザー評価のために2つのオンライン画像ラベリングシステムがReact.jsで開発された。(a) 非空間レイアウトラベリングシステムと(b) 空間レイアウトラベリングシステムである。

参加者内比較が採用され、各参加者は、2つのラベリングシステムを使用して2つのラベリングタスク（NSタスクとSタスク）をするように求められた。

・NSタスク：非空間レイアウトインターフェイスと条件

ベースライン条件となっている。図6(a)は、非空間レイアウトラベリングシステムのスクリーンショットである。インターフェイスの左側には、ラベルがテキストとサンプル画像の両方で一覧表示されている。インターフェイスの右側には、ラベル付けするターゲット画像が表示される。参加者は、20の犬種（20のラベル）のリストから適切なラベルを選択し、50の犬の画像にラベルを付けるように求められた。この非空間レイアウトインターフェイスでは、一度に1つの画像しか表示されない。参加者は、左側のリストをスクロールしてターゲット画像に適切なラベルを探し、暫定的な割り当てとしてラベルを選択した（選択したラベルがターゲット画像の下に表示された）。次に、「次へ」をクリックして割り当てを完了した。その後、選択したラベルについての自信を示すための「選択したラベルに自信があるか？」という質問に答えた。その後、「次へ」をクリックして次の画像に移動した。以前に選択したラベルを変更したい場合は、「前へ」をクリックして前の画像に戻ることも可能であった。

・Sタスク：空間レイアウトインターフェイスと条件
本研究の提案手法である。図6(b)は、空間レイアウトラベリングシステムのスクリーンショットを示している（空間レイアウトインターフェイスについてはセクション3で説明した）。ラベル付けタスク（50匹の犬の画像にラベルを付ける）には、非空間レイアウトインターフェイスに割り当てられたも

のと同じタスクが含まれていた。Sタスクでは、最終的なラベルを画像に割り当てる前に、最初にすべての画像と概念的な類似性を表すラベルをレイアウトするように参加者に指示した。

本ラベリングシステムは、参加者がラベリングタスクを完了するまでに要した時間とエラー率（画像に適切なラベルを選択できなかった）を自動的に記録および測定した。タイマーは、参加者が「開始」をクリックすると開始し、「終了」をクリックすると停止した。システムは、参加者がラベル付けに費やした時間も記録した。また、自信のあるラベル選択率（参加者がつけたラベルに自信を持っていた画像）も測定した。



(a) non-spatial layout system



(b) spatial layout system

図 6. 非空間および空間レイアウトラベリングシステムのスクリーンショット。

5 結果

図 7 は、参加者が平均 16 分 47 秒と 17 分 55 秒を費やして、非空間および空間レイアウトインターフェイスを使用して 50 枚の画像にそれぞれラベルを付けたことを示している。タスク完了時間に関して、対応のある t 検定の結果は、非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差がないことを示した ($p > 0.05$)。

図 8 は、非空間および空間レイアウトインターフェイスを使用して参加者が完了したラベリングタスクのエラー率（画像に適切なラベルを選択できない）を示している。結果として、エラー率が非空間レイアウトインターフェイスで 43.50%、空間レイアウト

インターフェイスで 37.63%であることがわかった。対応のある t 検定を使用したエラー率の分析では、2 つのラベリングインターフェイス間に有意差 ($p < 0.05$) があることが示された。

この結果は、非専門家のアノテーターが、タスク完了時間を大幅に増加させることなく、従来の非空間レイアウトインターフェイスよりも空間レイアウトインターフェイスを使用することで、特定のラベリングタスクをより正確に完了できることを示した。

画像データセットの 3 つの難易度レベルでエラー率を分析した (図 9 を参照)。結果は、レベル 1 とレベル 3 で非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差 ($p > 0.05$) がなかったのに対し、レベル 2 では有意差 ($p < 0.05$) があることがわかった。これは、空間レイアウトインターフェイスの利点が、ラベリングタスクにあいまいな画像が含まれている場合にのみ現れることを示している。この利点は、ターゲット画像があいまいでないか、あいまいすぎない場合には得られない。

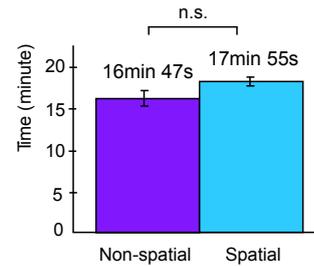


図 7. ラベル付けタスクの完了時間。

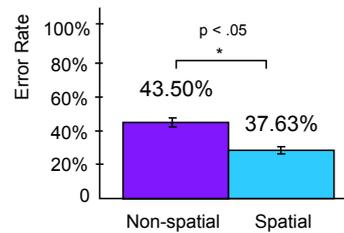


図 8. ラベリングタスクのエラー率。

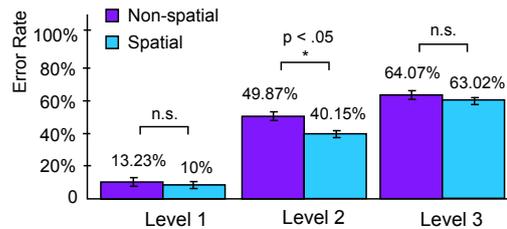


図 9. 難易度のエラー率。

ユーザー実験では、参加者は、選択したラベルの確信度（自信があるかどうか）を示すように求めら

れた。図 10 は、参加者が非空間レイアウトタスクでラベル付けされた画像の 47.13% (24/50 画像) に自信を持っていたのに対し、空間レイアウトタスクでは 59.63% (30/50 画像) であることを示している。対応のある t 検定の結果、非空間レイアウトインターフェイスと空間レイアウトインターフェイスの間に有意差 ($p < 0.01$) があることがわかった。これは、空間レイアウトインターフェイスが、手動の画像注釈に対する非専門家の主観的な自信を高めることができることを示している。

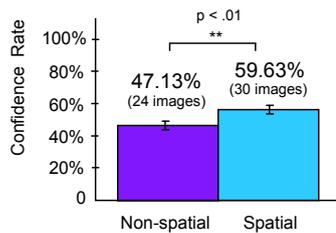


図 10. ラベリングタスクの信頼率。Non-spatial: mean = 47.13; SE = 4.52; Spatial: mean = 59.63; SE = 4.12; $p = 0.0008$.

6 今後の課題

本研究では、専門家でない画像注釈のラベル改善の本当の理由は明確には示されなかった。改善の理由が自己学習によるものなのか、注釈付け中に空間レイアウトインターフェイスを介して慎重にラベル付けするためなのかは不明である。将来的には、ラベルの改善の本当の理由（学習または注意深いラベル付け）を調査したいと考えている。

7 結論

本研究では、専門家以外の画像注釈のラベル品質を向上させるために、空間ラベリングと呼ばれる空間レイアウトラベリングインターフェイスを提案した。このインターフェイスは、ラベル付けプロセスでラベルを選択する前に画像とラベルをレイアウトするためのオープンスペース（つまり、空間編成プロセス）で構成されていた。本研究の空間レイアウトインターフェイスと従来の非空間レイアウトインターフェイスの比較のため、画像ラベリングタスクを用いてユーザー調査を実施した。結果は、非専門家のアノテーターは、従来の非空間レイアウトインターフェイスよりも本研究の空間レイアウトインターフェイスを使用した時に、画像のラベルをより正確に選択することを示した。さらに、空間レイアウトインターフェイスにより、手動の画像ラベリング中に専門家ではないアノテーターの確信度が向上することが観察された。空間ラベリングは、空間編成を介して非専門家の画像注釈のラベル品質を改善す

るための代替ソリューションを提供する。本研究の結果は、注釈ツールの将来の開発に使用できる重要な洞察を示している。

謝辞

本研究は JST CREST JPMJCR17A1 の支援を受けたものである。

参考文献

- [1] Abhishek Dutta, and Andrew Zisserman. 2019. The VIA Annotation Software for Images, Audio and Video. In Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276-2279.
- [2] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. 2008. LabelMe: A Database and Web-Based Tool for Image Annotation. International journal of computer vision 77, no. 1-3: 157-173.
- [3] Luis Von Ahn, and Laura Dabbish. 2004. Labeling Images with a Computer Game. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 319-326.
- [4] Chia-Ming Chang, Siddharth Deepak Mishra, and Takeo Igarashi. 2019. A Hierarchical Task Assignment for Manual Image Labeling. In 2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 139-143.
- [5] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2334-2346.
- [6] Takeru Sunahase, Yukino Baba, and Hisashi Kashima. 2017. Pairwise HITS: Quality Estimation from Pairwise Comparisons in Creator-Evaluator Crowdsourcing Process. In Thirty-First AAAI Conference on Artificial Intelligence.
- [7] Yi-Li Fang, Hai-Long Sun, Peng-Peng Chen, and Ting Deng. 2017. Improving the Quality of Crowdsourced Image Labeling via Label Similarity. Journal of Computer Science and Technology 32, no. 5: 877-889.
- [8] Todd Kulesza, Saleema Amershi, Rich Caruana, Danyel Fisher, and Denis Charles. 2014. Structured Labeling for Facilitating Concept Evolution in Machine Learning. In Proceedings of the SIGCHI Conference on Human Factors in

- Computing Systems. ACM, 3075–3084.
- [9] Ken Perlin, and David Fox. 1993. Pad: An Alternative Approach to the Computer Interface. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques, pp. 57-64.
- [10] Benjamin B. Bederson, and James D. Hollan. 1994. Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. In Proceedings of the 7th annual ACM symposium on User interface software and technology, pp. 17-26.
- [11] Daniel Bauer, Pierre Fastrez, and Jim Hollan. 2005. Spatial Tools for Managing Personal Information Collections. In Proceedings of the 33th IEEE Annual Hawaii International Conference on System Sciences, pp. 104b-104b.
- [12] Nayuko Watanabe, Motoi Washida, and Takeo Igarashi. 2007. Bubble Clusters: An Interface for Manipulating Spatial Aggregation of Graphical Objects. In Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 173-182.
- [13] George Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten Van Dantzich. 1998. Data Mountain: Using Spatial Memory for Document Management. In Proceedings of the 11th annual ACM symposium on User interface software and technology, pp. 153-162.
- [14] Nan-Chen Chen, Jina Suh, Johan Verwey, Gonzalo Ramos, Steven Drucker, and Patrice Simard. 2018. AnchorViz: Facilitating Classifier Error Discovery through Interactive Semantic Data Exploration. In 23rd International Conference on Intelligent User Interfaces, pp. 269-280.
- [15] Yanir Kleiman, Joel Lanir, Dov Danon, Yasmin Felberbaum, and Daniel Cohen-Or. 2015. DynamicMaps: Similarity-based Browsing through a Massive Set of Images. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 995-1004.
- [16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. 2015. ImageNet Large Scale Visual Recognition Challenge. International journal of computer vision 115, no. 3: 211-252.
- [17] Chia-Ming Chang, Chia-Hsien Lee, and Takeo Igarashi. 2021. Spatial Labeling: Leveraging Spatial Layout for Improving Label Quality in Non-Expert Image Annotation. In Proceedings of the ACM Conference on Human Factors in Computing Systems. ACM.

未来ビジョン

通常、高品質のデータは専門家のアノテーターに依存している。ただし、優れた注釈ツールが提供されていれば、専門家でない注釈者も高品質のデータを生成できると考えている。この研究は、ラベリングインターフェースの設計が非専門家のデータ注釈のラベル品質を改善できるという成功例を示している。今日、AIの時代では、新しいインテリジェントシステムには常に膨大な数のトレーニングデータが必要である。専門家ではないアノテーターがアノテーション中にドメイン知識を構築するのに役立つアノテーションツールを提供できれば、それは大きな貢献になる可能性がある。将来的には、難しいアノテーション作業でも誰もがアノテーターになれるように願っている。