

密度に基づく時空間分析システムにおける 学習済み深層ネットワークを用いた画像分類

Image Classification using Pre-trained Deep Network for Density-based Spatiotemporal Analysis System

酒井 達弘^{1,2*} 田村 慶一¹ 北上 始¹
Tatsuhiro Sakai^{1,2} Keichi Tamura¹ Hajime Kitakami¹

¹ 広島市立大学大学院情報科学研究科

¹ Graduate School of Information Sciences, Hiroshima City University

² 日本学術振興会特別研究員 DC1

² JSPS Research Fellow DC1

Abstract: Recently, during natural disasters like, earthquakes, typhoons, flood, and heavy snowfall, people actively post messages that mention situations people are facing through social media sites. Therefore, the enhancement of situation awareness in the real world using social data is one of the most attractive research topics. In our previous work, we developed a density-based spatiotemporal analysis system to identify topic-related areas in which there are a huge number of geo-tagged tweets related to a topic are posted. In this paper, we propose a novel density-based spatiotemporal analysis system with a photo image classifier using the pre-trained deep network in order to enhance situation awareness by showing accurate topic-related photos. The photo image classifier using a support vector machine (SVM) based on the pre-trained deep network is integrated into the conventional density-based spatiotemporal analysis system. To evaluate the proposed system, we used actual tweet data sets related to weather topics, “heavy rain” and “heavy snow,” in Japan. The experimental results showed that the proposed system can classify photo images related to these weather topics more sensitively compared with our previous method.

1 はじめに

近年、インターネット上のユーザはソーシャルメディアサイトを通して気象状況や発生した自然災害に関する内容を投稿するようになってきている。例えば、代表的なソーシャルメディアサイトの Twitter では、大雨、大雪、地震や台風などを目の当たりにした人がテキストメッセージや写真によってそれらの状況を伝えている [1]。通常、このような投稿には位置情報（ジオタグと呼ぶ）が付与されており、これらのジオタグ付きのツイートから気象状況や自然災害などの緊急性のあるトピックを検出し、活用することが期待されている [2]。

そこで我々は、気象状況や自然災害などモニタリングをしたいトピックを設定し、当該トピックの発生を検出し、時空間上での変化を分析するために密度に基

づく時空間分析システムを提案している [3, 4]。密度に基づく時空間分析システムでは、ナイーブベイズ分類器、 (ϵ, τ) -密度に基づく適応的な時空間クラスタリングと Bag-of-Features (BoF) の 3 つの手法を用いている。ナイーブベイズ分類器を用いてトピックに関連するツイートのみを抽出する。 (ϵ, τ) -密度に基づく適応的な時空間クラスタリングを用いることで、トピックに関連するジオタグ付きツイートが時空間上で密集している領域を時空間クラスタとして抽出し、トピックの発生や消滅を監視することができるさらに、BoF を用いて時空間クラスタに含まれる画像データから特徴ベクトルを抽出し、サポートベクターマシーン (SVM) によってトピックに関連する画像データかどうか分類し、トピックに関連する画像データのみを抽出する。また、トピックの時空間上での変化を確認するための Web インタフェースから、抽出された時空間クラスタと画像データの情報を閲覧することができる。

先行研究では、画像データの特徴ベクトル抽出器として BoF を用いていたが、一定の精度でトピックに関

*連絡先：広島市立大学大学院情報科学研究科
〒731-3194 広島市安佐南区大塚東 3-4-1
E-mail: da65003@e.hiroshima-cu.ac.jp

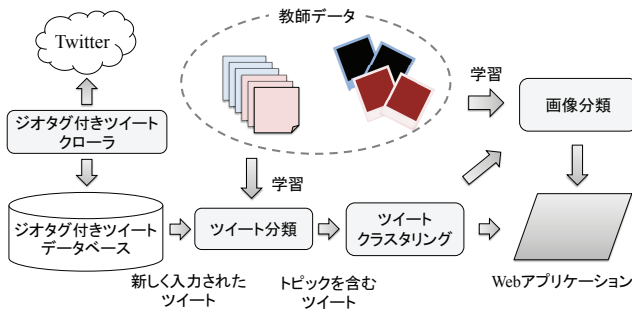


図 1: 密度に基づく時空間分析システムの概要

連する画像データの分類が可能であるものの、分類精度の向上が課題となっていた。そこで本研究では、密度に基づく時空間分析システムにおける画像分類手法として、学習済み深層ネットワークを用いた画像分類を導入する。大規模画像データを用いて学習させた深層ネットワークから抽出した特徴ベクトルは、BoFを用いて抽出した特徴ベクトルよりも汎用性が高く、画像データの特徴を十分に捉えることができ、高精度に分類することができる。

本論文の構成は以下の通りである。第2章では、密度に基づく時空間分析システムについて、簡単に説明する。第3章では、提案手法である学習済み深層ネットワークを用いた画像分類を導入したシステムを説明する。第4章では、評価実験の結果を示し、第5章で本論文をまとめる。

2 密度に基づく時空間分析システム

本章では、密度に基づく時空間分析システムについて簡単に説明する。

2.1 データ定義

Twitter 上から取得したジオタグ付きツイートを gt_i と表記し、その集合を $GTS = \{gt_1, \dots, gt_n\}$ とする。ここで、 gt_i は文書データ $text_i$ 、投稿時間 pt_i 、位置情報 pl_i と画像データ pi_i の4つから構成される。本研究では、位置情報として経度と緯度を用いる。また、モニタリングをしているトピックの内容を含むジオタグ付きツイートを関連ジオタグ付きツイート $rgt_j (= gt_{\phi(j)})$ と呼ぶ。関連ジオタグ付きツイート集合を $RGTS = \{rgt_1, \dots, rgt_m\}$ とすると、 GTS は $RGTS$ を包含しており ($RGTS \subset GTS$)、次の単射で表現される。

$$\phi(j) : RGTS \rightarrow GTS; rgt_j \mapsto gt_{\phi(j)} \quad (1)$$



図 2: Web アプリケーション画面

例えば、“大雨”についてモニタリングをしているとすると、関連ジオタグ付きツイートとは、“すごい大雨”や“強い雨が降り始めた”などの大雨が降っていることを伝えているツイートとなる。

2.2 システム概要

図1に密度に基づく時空間分析システムの概要図を示す。密度に基づく時空間分析システムでは、ツイート分類、ツイートクラスタリングと画像分類を一定時間毎に実行し、Web アプリケーション上へデータを出力する。システム全体の処理の流れを次に示す。

1. ジオタグ付きツイートクローラを用いて Twitter からジオタグ付きツイートを収集し、ジオタグ付きツイートデータベースに保存する。
2. ツイート分類部では、収集したジオタグ付きツイートに対して、ナイーブベイズ分類器を用いてモニタリングをしているトピックに関連する関連ジオタグ付きツイートとそれ以外のツイートに分類を行い、関連ジオタグ付きツイートのみを抽出する。
3. ツイートクラスタリング部では、新たに収集された関連ジオタグ付きツイートとこれまでに抽出された時空間クラスタ集合を入力として、 (ϵ, τ) -密度に基づく適応的な時空間クラスタリング [5] を用いて新しく時空間クラスタ集合を抽出する。新たに時空間クラスタリングが抽出されるとトピックの発生、時空間クラスタが消滅すると時空間ク

ラスタが存在した地域においてトピックが消滅したことを意味する。

4. 画像分類部では、時空間クラスタに含まれる関連ジオタグ付きツイートから画像データを取り出し、BoF[6]を用いて各画像データの特徴ベクトルを抽出する。次に、SVMを用いてモニタリングをしているトピックに関連している関連画像データとそれ以外の画像データに分類を行い、関連画像データのみを抽出する。SVMは事前に教師データを用いて学習済みであるとする。
5. 時空間クラスタの内容と関連画像データをWebアプリケーション上へ出力する。図2にWebアプリケーションのスクリーンショットを示す。Webアプリケーションでは、地図上に時空間クラスタとして抽出されたツイートと関連画像データが表示される。

3 提案システム

本章では、学習済み深層ネットワークを用いた画像分類を導入した新しい密度に基づく時空間分析システムを提案する。

3.1 概要

提案システムでは、先行システムにおける画像分類について、学習済み深層ネットワークを特徴ベクトル抽出器として利用する。先行システムの特徴ベクトル抽出器としてBoFを用いた画像分類は、一定の精度でトピックに関連する画像データの分類が可能であるものの、その精度の向上が課題となっていた。Twitterに投稿される画像データはバリエーションが多く、決まった種類の画像データが存在しないために、BoFでは画像データの特徴を十分に捉えることができない。提案システムでは、学習済みの畳み込みニューラルネットワーク(CNN)を画像データの特徴ベクトル抽出器として用いる。

大規模データセットを用いて学習させたCNNの中間層から抽出される特徴ベクトルは非常に汎用性が高く、様々なドメインで再利用可能であることが示されている。教師データとなる画像データ集合について学習済みの畳み込みニューラルネットワークから特徴ベクトルを抽出する。次に、SVMを用いて再学習し、分類器を構成する。

表 1: VGG-16 の構造

層名	ユニット数
入力層	150528
第 1 層 (畳み込み層)	64
第 2 層 (畳み込み層)	64
プーリング層	
第 3 層 (畳み込み層)	128
第 4 層 (畳み込み層)	128
プーリング層	
第 5 層 (畳み込み層)	256
第 6 層 (畳み込み層)	256
第 7 層 (畳み込み層)	256
プーリング層	
第 8 層 (畳み込み層)	512
第 9 層 (畳み込み層)	512
第 10 層 (畳み込み層)	512
プーリング層	
第 11 層 (畳み込み層)	512
第 12 層 (畳み込み層)	512
第 13 層 (畳み込み層)	512
プーリング層	
第 14 層 (全結合層)	4096
第 15 層 (全結合層)	4096
第 16 層 (出力層)	1000

3.2 学習済み深層ネットワークを用いた特徴ベクトル抽出

提案システムで用いるCNNは、数多く提案されている深層ネットワークの中でも、特に画像認識の分野に応用されているニューラルネットワークである。CNNは中間層に畳み込み層とプーリング層が存在し、画像データの局所的な特徴を自動的に学習することができる。

提案システムでは、特徴ベクトル抽出器としてVGG-16と呼ばれるCNNのネットワークモデルを用いる[7]。VGG-16の学習は、大規模画像認識コンペティションのILSVRC-2014で提供されたImageNetの1000分類、120万枚の画像を用いて行われている。ImageNetは一般的な内容の画像データを含んでおり、VGG-16は画像データの様々な分析に応用可能な汎用知識を学習できていると言われている。

VGG-16の具体的な構造を表1に示す。VGG-16は16層から形成される。畳み込み層では、畳み込み処理を行っており、入力に対して重みフィルタの内積を計算する。各畳み込み層は前層の入力に対して畳み込み処理を行い、次の層の入力となる特徴マップを出力する。プーリング層では、畳み込み層から出力された特徴マップを縮小する。VGG-16では、最大値プーリン

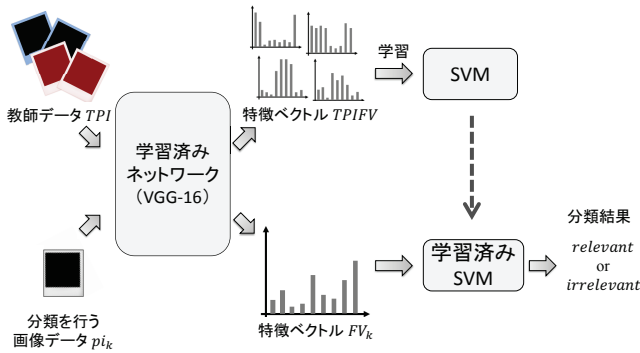


図 3: 提案システムの画像分類の処理手順

グを用いている。全結合層では、重み付き結合を計算し、活性化関数によりユニットの値を求める。VGG-16 では活性化関数として、ReLU を用いている。

提案システムでは、第 15 層の全結合層から特徴ベクトルを抽出するため、各画像データについて 4096 次元の特徴ベクトルが抽出される。画像データ i の特徴ベクトルを、 $FV_i = \{fv_{i,1}, fv_{i,2}, \dots, fv_{i,4096}\}$ とする。VGG-16 の学習に用いられている ImageNet には気象や自然災害に関する分類を含む画像データは無い。しかしながら、出力層手前の中間層には画像データの汎用的な特徴が表れるため、ソーシャルメディア上に投稿される画像データを区別する特徴ベクトルとして利用できると考える。

3.3 処理手順

提案システムは学習済み深層ネットワークを用いて抽出した特徴ベクトルと SVM を用いて画像分類を行う。教師データを $TPI = \{(tpi_1, tc_1), (tpi_2, tc_2), \dots, (tpi_l, tc_l)\}$ とし、 $tc_i = \{relevant, irrelevant\} \in CLASS$ とする。 $relevant$ はモニタリングをしているトピックに関連している画像データのクラスとし、 $irrelevant$ はそれ以外の画像データのクラスである。提案システムの画像分類の処理手順を図 3 に示す。処理手順は次の通りである。

1. 教師データ TPI の各画像データから学習済み深層ネットワークを用いて特徴ベクトル $TPI_FV = \{FV_1, FV_2, \dots, FV_n\}$ を抽出する。
2. TPI_FV を用いて SVM を学習させる。
3. 分類を行う画像データ p_k から学習済み深層ネットワークを用いて特徴ベクトル FV_k を抽出し、SVM で分類を行う。 $relevant$ クラスに分類された画像データを関連画像データとする。

4 評価実験

提案手法の有効性を確認するために、評価実験を行った。本章では、評価実験の結果を示す。

4.1 実験内容

評価実験では、モニタリングをするトピックを“大雨”と“大雪”としてそれぞれ評価を行う。比較手法としては、画像データの特徴ベクトル抽出器として、先行システムで用いられていた BoF を用いた手法 (BoF と表記する)、VGG-16 と同じ構造をした 16 層の CNN を構築し教師データを用いて学習させたネットワークを用いた手法 (CNN-16 と表記する)、提案手法である VGG-16 のネットワークを用いた手法 (VGG-16 と表記する) とで比較を行う。BoF によって抽出する特徴ベクトルの次元数は 4096 と設定した。

評価方法としては、最初に作成した教師データについて交差検定によって評価を行う。次に、実際に密度に基づく時空間分析システムによって抽出された時空間クラスタから画像データを取り出し、画像分類を行った結果を評価する。教師データ TPI は、トピック“大雨”については 2014 年 7 月に投稿された $relevant$ クラス 500 件と $irrelevant$ クラス 500 件を用いた。また、トピック“大雪”については 2014 年 1 月と 2 月に投稿された $relevant$ クラス 500 件と $irrelevant$ クラス 500 件を用いた。

4.2 交差検定

最初に交差検定による評価を行う。交差検定の分割数は 2, 4, 6, 8 と 10 分割を用いた。モニタリングをしているトピックを“大雨”としたときの交差検定の結果を図 4 に、モニタリングをしているトピックを“大雪”としたときの交差検定の結果を図 5 に示す。図 4 と図 5 には、正解率、精度と再現率をそれぞれ示している。図 4 と図 5 より、正解率、精度と再現率の全てについて、VGG-16 は BoF と CNN-16 よりも高性能であることを示した。

4.3 システム評価

次に、実際に抽出された時空間クラスタに含まれているジオタグ付きツイートに付与されている画像データを分類し、評価を行う。実験期間としては、トピック“大雨”については、日本で台風が観測され全国各地で大雨となった 2014 年 8 月 1 日から 10 日としている。トピック“大雪”については、日本全国各地で降雪のあった 2014 年 2 月 10 日から 2 月 18 日とする。

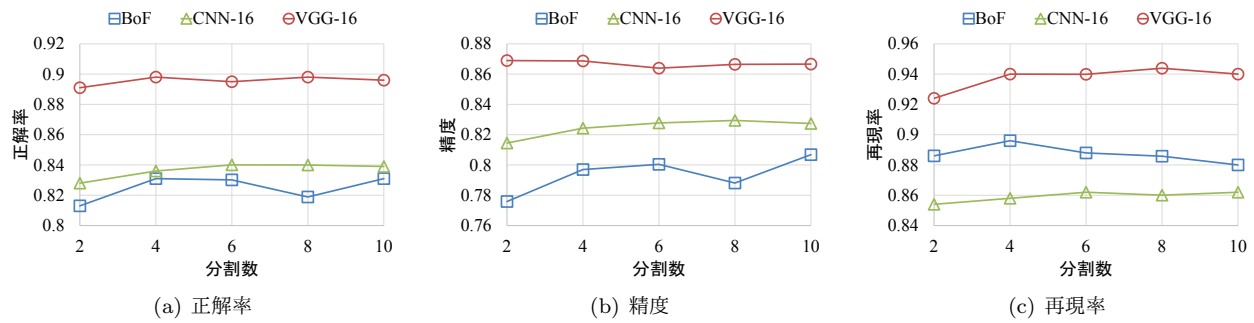


図 4: トピック“大雨”の交差検定の結果

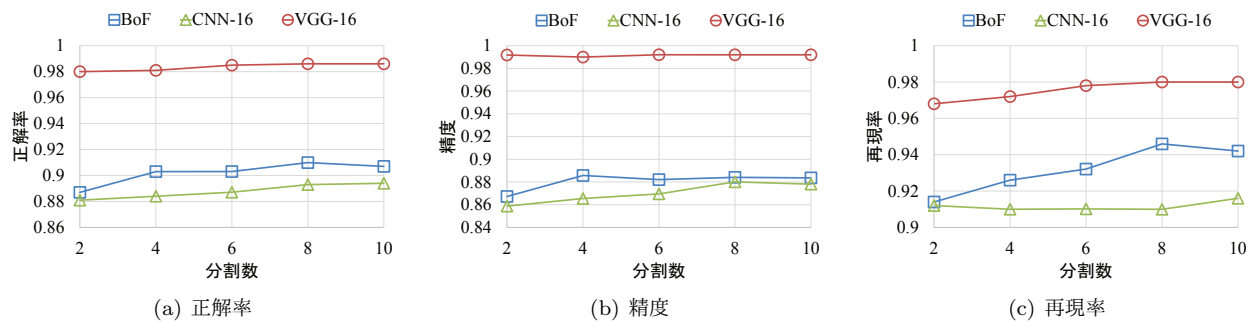


図 5: トピック“大雪”の交差検定の結果

モニタリングをしているトピックを“大雨”としたときの実験結果を表2に、モニタリングをしているトピックを“大雪”としたときの実験結果を表3に示す。表2と表3には、各日付の時空間クラスに含まれていた画像データ数、正解率、精度と再現率をそれぞれ示している。表2より、トピック“大雨”について正解率を見ると、VGG-16は10日間で最も高い正解率を示している。精度と再現率についてもVGG-16が最も良い結果を示している。また、表3より、トピック“大雪”についても同様に、VGG-16が最も良い結果が得られた。図6と図7に、BoFでは正しく *relevant* クラスに分類されなかったが、VGG-16では正しく *relevant* クラスへ分類することができた画像データの例を示す。VGG-16によって正しく *relevant* クラスへ分類された画像データには、モニタリングをしているトピック以外の人物や物体が写っているものが多いことが分かる。VGG-16を用いて抽出された特徴ベクトルは、モニタリングをしているトピック以外のものが写っていたとしても、画像データ中の当該トピックの部分の特徴を捉えることができたと言える。以上の実験結果より、大規模画像データによって学習済み深層ネットワークを特徴ベクトル抽出器として用いることの有効性を示すことができた。

5 まとめ

本研究では、画像分類として学習済み深層ネットワークを用いた画像分類を導入した新しい密度に基づく時空間分析システムを提案した。大規模画像データを用いて学習させた深層ネットワークから抽出した特徴ベクトルは、BoFを用いて抽出した特徴ベクトルよりも汎用性が高く、画像データの特徴を十分に捉えることができ、高精度に分類することができる。評価実験より、提案手法は先行研究の手法よりも高性能に画像分類ができることを示した。今後の課題としては、学習済み深層ネットワークを再学習させて新しいモデルを作成することがあげられる。学習済みの深層ネットワークを初期値とし再学習することで、汎用性があり、さらに特定の内容に最適な深層ネットワークができることを期待できる。

謝辞

本研究の一部は、JSPS 科研費 JP16J05403, JP26330139 と総務省 SCOPE (受付番号:162308002) の支援により行われた。

表 2: トピック “大雨” のシステム評価

日付	画像データ数	正解率			精度			再現率		
		BoF	CNN-16	VGG-16	BoF	CNN-16	VGG-16	BoF	CNN-16	VGG-16
8/1	130	0.68	0.73	<u>0.85</u>	0.59	0.66	<u>0.75</u>	0.76	0.76	<u>0.95</u>
8/2	217	0.76	0.71	<u>0.83</u>	0.54	0.49	<u>0.63</u>	<u>0.97</u>	0.82	0.92
8/3	203	0.75	0.78	<u>0.82</u>	0.60	0.67	<u>0.69</u>	<u>0.86</u>	0.73	<u>0.86</u>
8/4	63	<u>0.83</u>	0.67	0.81	<u>0.57</u>	0.37	0.54	<u>0.93</u>	0.71	<u>0.93</u>
8/5	63	0.67	0.70	<u>0.78</u>	0.43	0.45	<u>0.53</u>	0.94	0.88	<u>1.00</u>
8/6	87	0.61	0.63	<u>0.64</u>	0.34	0.36	<u>0.38</u>	0.84	0.89	<u>0.95</u>
8/7	53	0.58	0.58	<u>0.70</u>	0.39	0.37	<u>0.47</u>	<u>1.00</u>	0.79	<u>1.00</u>
8/8	230	0.62	0.71	<u>0.74</u>	0.34	0.41	<u>0.44</u>	0.82	0.80	<u>0.90</u>
8/9	412	0.70	0.70	<u>0.77</u>	0.39	0.38	<u>0.46</u>	<u>0.90</u>	0.82	0.87
8/10	572	0.73	0.74	<u>0.81</u>	0.56	0.58	<u>0.66</u>	0.88	0.81	<u>0.90</u>

表 3: トピック “大雪” のシステム評価

日付	画像データ数	正解率			精度			再現率		
		BoF	CNN-16	VGG-16	BoF	CNN-16	VGG-16	BoF	CNN-16	VGG-16
2/10	210	0.68	0.68	<u>0.80</u>	0.74	0.75	<u>0.83</u>	0.73	0.71	<u>0.84</u>
2/11	206	0.70	0.71	<u>0.81</u>	0.74	0.76	<u>0.81</u>	0.82	0.80	<u>0.92</u>
2/12	23	0.65	<u>0.70</u>	<u>0.70</u>	0.29	<u>0.38</u>	<u>0.38</u>	0.40	<u>0.60</u>	<u>0.60</u>
2/13	87	0.67	0.68	<u>0.77</u>	0.61	0.63	<u>0.71</u>	0.75	0.73	<u>0.85</u>
2/14	2196	0.78	0.77	<u>0.87</u>	0.82	0.84	<u>0.89</u>	0.86	0.80	<u>0.93</u>
2/15	940	0.75	0.81	<u>0.90</u>	0.86	0.90	<u>0.93</u>	0.79	0.84	<u>0.93</u>
2/16	306	0.69	0.71	<u>0.85</u>	0.72	0.76	<u>0.82</u>	0.80	0.76	<u>0.96</u>
2/17	152	0.68	0.69	<u>0.80</u>	0.66	0.71	<u>0.73</u>	0.77	0.68	<u>0.95</u>
2/18	58	0.57	<u>0.66</u>	<u>0.66</u>	0.44	<u>0.52</u>	0.51	0.71	0.62	<u>0.90</u>



図 6: VGG-16 によって正しく *relevant* クラスへ分類された画像データの例 (トピック “大雨”)



図 7: VGG-16 によって正しく *relevant* クラスへ分類された画像データの例 (トピック “大雪”)

参考文献

- [1] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: Can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics, SOMA '10*, pp. 71–79, 2010.
- [2] Mor Naaman. Geographic information from geo-referenced social media data. *SIGSPATIAL Special*, Vol. 3, No. 2, pp. 54–61, 2011.
- [3] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami. Density-based spatiotemporal analysis system with photo image classifier using the BoF model. *Information Engineering Express*, Vol. 1, No. 4, 2015.
- [4] Tatsuhiro Sakai and Keiichi Tamura. Real-time analysis application for identifying bursty local areas related to emergency topics. *SpringerPlus*, Vol. 4, No. 162, 2015.
- [5] Tatsuhiro Sakai, Keiichi Tamura, and Hajime Kitakami. Emergency situation awareness during natural disasters using density-based adaptive spatiotemporal clustering. In *Database Systems for Advanced Applications, DASFAA 2015 International Workshops, SeCoP, BDMS, and Posters, Hanoi, Vietnam, April 20-23, 2015*, Vol. 9052, pp. 155–169, 2015.
- [6] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cedric Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.
- [7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, arXiv:1409.1556, 2014.