ImageNet Learge Scale Visual Recognition Challenge

備考

ILSVRC 2010-2014 までの優勝モデル

	Classification(分類)		**Localization(定位)**		**Detection(検出)**	
	提供されたデータ	外部データ	提供されたデータ	外部データ	提供されたデータ	外部データ
2010	確率的SVM (NEC)	-	-	-	-	-
2011	1対他 線形SVM	-	ヒストグラム交差カーネルSVM	-	-	-
2012	AlexNet	-	AlexNet	-	-	-
2013	Clarifai	-	FCN	-	Over Feat	-
2014	GoogLeNet	弱教師付き学習物体検出器 CASIAWS	VGGNet	VGGNet	R-CNN + Network-in-Network	GoogLeNet

著者

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei

掲載

International Journal of Computer Vision (IJCV), Vol. 115, pp. 211-252, 2015.

Abstract

ImageNet 大規模視覚認識チャレンジは、数百のオブジェクトカテゴリと数百万の画像を対象としたオブジェクトカテゴリの分類と検出のベンチマークです。このチャレンジは2010年から現在まで毎年実施されており、50以上の機関から参加しています。

本論文では、このベンチマークデータセットの作成と、その結果として可能となった物体認識の進歩について述べる。 大規模な基底真実アノテーションを収集する上での課題を議論し、分類的物体認識における重要なブレークスルーを強調し、大規模な画像分類と物体検出の分野の現状を詳細に分析し、最先端のコンピュータビジョンの精度と人間の精度を比較する。 最後に、5年間の研究成果から得られた教訓を紹介し、今後の方向性と改善点を提案する。

Introduction

1ILSVRCは、2005年に設立された「PASCAL VOCチャレンジ」（PASCAL VOC）の流れを汲むもので、毎年のコンペティション形式で認識アルゴリズムの標準的評価を行ってきました。 ILSVRCは、PASCAL VOCと同様、以下の2つの要素から構成されています。 (1)一般に公開されているデータセット (2)年次大会とそれに対応するワークショップ データセットでは、分類的な物体認識アルゴリズムの開発と比較を行うことができ、コンペとワークショップでは、毎年、最も成功した革新的なエントリーから得られた教訓について、進捗状況を追跡し、議論することができます。

公開されているデータセットには、手動で注釈を付けた訓練画像セットが含まれています。参加者は訓練画像を用いてアルゴリズムの訓練を行い、その訓練画像に自動的に注釈を付けます。予測されたアノテーションは評価サーバに送信されます。評価結果は期間終了後に発表され、国際コンピュータビジョン会議（ICCV）または欧州コンピュータビジョン会議（ECCV）で開催されるワークショップで発表されます。

ILSVRC のアノテーションは 2 つのカテゴリに分類されます： (1)画像レベルのアノテーションでは、画像中のオブジェクトクラスの有無を示すバイナリラベル（例：「この画像には車がある」が「虎はいない」） (2)オブジェクトレベルのアノテーションでは、画像中のオブジェクトインスタンスの周りにある狭い境界ボックスとクラスラベル（例：「幅 50 ピクセル、高さ 30 ピクセルの位置(20,25)にドライバーがあります」） です。

大規模なチャレンジとイノベーション

データセットを作成するにあたり、いくつかの課題があった。 PASCAL VOC 2010の19,737画像からILSVRC 2010の1,461,406枚の画像へ、またオブジェクトクラスを20から1,000へとスケールアップするには、いくつかの課題がある。他のデータセットで行われているように、少数のアノテータがデータのアノテーションを行うことはもはや不可能である(Fei-Feiet al., 2004; Criminisi, 2004; Everingham et al., 2012; Xiao et al., 2010)。その代わりに、我々は大規模なアノテーションを収集するための新しいクラウドソーシングのアプローチを設計することに目を向ける(Su et al., 2012; Deng et al., 2009, 2014)。

例えば、束になっているバナナは、飛行機や自動車のような基本レベルのカテゴリに比べて、アノテーションが容易ではないかもしれません。また、100万枚以上の画像があるため、すべての対象物の位置をアノテーションすることは不可能である（PASCAL VOCのサブセットに含まれる対象物のセグメンテーションや人体部分、その他の詳細なアノテーションでは不可能である）。このような状況では、完全な人間の手によるアノテーションを得ることができない可能性があることを考慮して、新たな評価基準を定義する必要がある。

挑戦的なデータセットが収集されると、その規模の大きさから、物体認識アルゴリズムの評価と新しい技術の開発の両方において、これまでにないチャンスが生まれました。大規模な学習データが利用可能になると、新しいアルゴリズムの革新が生まれます。 対象物のカテゴリが多岐にわたるため、視覚的に非常に類似したクラスを識別できるアルゴリズムが必要とされています。本論文では、これらのアルゴリズムの中で最も成功したアルゴリズムを取り上げ、その性能を人間レベルの精度と比較する。

最後に、ILSVRCに収録されているオブジェクトクラスの種類が多いため、オブジェクトの統計的特性と認識アルゴリズムへの影響を解析する。

本論文では、3つの主要な目標を掲げています。

1. この大規模物体認識ベンチマークデータセットを作成する際の課題を議論する。

2. この成果から得られた物体の分類と検出の発展を強調すること。
3. 本論文は、大規模データセットの作成に携わる研究者だけでなく、大規模物体認識の歴史と現状をより深く理解したいと考えている方にも興味を持っていただけると思います。

1.1. Related work

ベンチマーク画像データセットの構築に関する先行研究について簡単に述べる。

画像分類データセット.Caltech 101 (Fei-Fei et al.,2004)は、最初に標準化されたマルチカテゴリ画像分類のデータセットの一つで、101のオブジェクトクラスと一般的にクラスごとに15〜30のトレーニング画像を持っていました。Caltech256 (Griffin et al., 2007)は、オブジェクトクラスの数を256に増やし、スケールと背景のばらつきを大きくした画像を追加しました。 別のデータセットTinyIm-ages (Torralba et al., 2008)には、インターネットから収集した8,000万枚の32x32の低解像度画像が含まれており、WordNet (Miller, 1995)のシンセットをクエリとして使用している。しかし、このデータは手動で検証していないため、エラーが多く、アルゴリズムの評価には不向きである。

ILSVRC は ImageNet データセット (Deng et al., 2009) がバックボーンとなっている。 ImageNetは、WordNet階層（Miller,1995）に従って構成された画像データセットである。WordNetの各概念は、複数の単語や語句で記述されている可能性があり、「同義語セット」または「同義語セット」と呼ばれています。ImageNetは、平均650個の手動検証済みの完全解像度の画像を用いて、21,841個のシンセットを生成している。その結果、ImageNetには14,197,122枚の注釈付き画像が含まれており、WordNetのセマンチエラルキーに基づいて整理されています（2014年8月現在）。ImageNetは、他の画像クラス分類データセットよりも規模が大きく、多様性に富んでいます。ILSVRCは、アルゴリズムのトレーニングにImageNet画像のサブセットを使用し、アルゴリズムをテストするための追加画像のアノテーションにImageNetの画像収集プロトコルの一部を使用しています。

画像解析データセット

いくつかのデータセットは、画像カテゴリーラベルを超えて、より豊富な画像アノテーションを提供することを目的としています。LabelMe (Russell et al., 2007)は、画像ごとに複数のオブジェクトを持つ一般的な写真を収録しています。 LabelMeは、オブジェクトの周りに多角形の注釈を付けていますが、ほとんどの部分が完全にラベル付けされておらず、オブジェクトの名前も標準化されていません。このため、LabelMeをアルゴリズムの学習や評価に利用することは困難です。SUN2012 (Xiao et al.,2010)のデータセットには、オブジェクト検出に適した16,873枚の画像が含まれています。LotusHill (Yao et al.,2007)のデータセットには、636,748枚の画像とビデオフレームのオブジェクトの詳細なアノテーションが含まれていますが、無料で利用することはできません。 ピクセルレベルのセグメンテーションを提供しているデータセットがいくつかあります。例えば、591枚の画像と23のオブジェクトクラスを持つMSRCデータセット(Criminisi, 2004)、715枚の画像と8つのクラスを持つStanford Background Dataset(Gouldet al., 2009)、500枚の画像にオブジェクトの境界をアノテーションしたBerkeley Segmentationデータセット(Arbelaez et al., 2011)などがあります。

ILSVRCに最も近いのはPASCAL VOCデータセット(Everingham et al., 2010, 2014)であり、物体検出、画像分類、物体セグメンテーション、人物レイアウト、行動分類のための標準化されたテストベッドを提供しています。 ILSVRの設計選択の多くはPASCAL VOCにインスパイアされており、データセット間の共通点や相違点については本稿で詳しく説明しています。ILSVRCは、PASCAL VOCが目標としていた認識アルゴリズムの標準化された学習と評価を、対象物のクラス数と画像数で一桁以上スケールアップしたものである。 PASCAL VOC 2012の物体クラス数は20、画像数は21,738枚であるのに対し、ILSVRC2012の物体クラス数は1,000、注釈付き画像数は1,431,167枚である。

最近リリースされたCOCOデータセット(Lin et al.,2014b)には、手動でセグメント化された250万個の物体インスタンスを含む328,000枚以上の画像が含まれています。ILSVRCに比べてオブジェクトのカテゴリ数は少ないが（COCOでは91個、ILSVRCのオブジェクト検出では200個）、カテゴリあたりのインスタンス数は多い（ILSVRCのオブジェクト検出では約1Kであったのに対し、平均で27K）。さらに、現在ILSVRCでは利用できないオブジェクトセグメンテーションアノテーションが含まれています。COCOは、今後も重要な大規模ベンチマークとなる可能性があります。

大規模アノテーション

ILSVRC は、正確なアノテーションを得るために Amazon Mechanical Turk を広く利用している（Sorokin and Forsyth, 2008）。(Welin-der et al., 2010; Sheng et al., 2008; Vittayakorn andHays, 2011)などの著作では、マーケットプレイスの品質管理メカニズムが記述されている。(Vondrick et al., 2012)は、クラウドソーシングのビデオアノテーションのデテールの概要を提供しています。 クラウドソーシングによる正確な画像アノテーションへの新しいアプローチについては、3.1.3項、3.2.1項、3.3.3項を参照されたい。

標準化された課題

ILSVRCと同様のオンライン評価に耐えるデータセットがいくつかあります：前述のPASCAL VOC (Everingham et al., 2012)、制約のない顔認識のためのLabeled Faces in the Wild (Huang et al., 2007)、3D再構成のための Reconstruction meetsRecognition (Urtasun et al., 2014)、自律運転におけるコンピュータビジョンのための KITTI (Geiger et al., 2013)などです。これらのデータセットとILSVRCは、コンピュータビジョンのさまざまな分野での進歩をベンチマークするのに役立ちます。

1.2. Paper layout

第2節でILSVRCの課題の概要を簡単に説明します。 第3節では、データセットの収集とアノテーションについて詳しく述べる。第4節では、大規模認識設定におけるアルゴリズムの評価基準について議論する。第5節では、ILSVRC参加者が開発した手法の概要を紹介する。

第6節では、ILSVRCの結果の詳細な分析を行う。 第6.1節では長年にわたる大規模認識の進展について、第6.2節ではILSVRCの結果は統計的に有意であると結論づけ、第6.3節では物体認識の分野の現状を徹底的に分析し、第6.4節では最新のコンピュータビジョンの精度と人間の精度を比較しています。最後に、第7節でILSVRCから得られた教訓について述べる。

4. Evaluation at large scale

データセットを収集した後は、アルゴリズムの標準化された評価手順を定義する必要がある。画像分類のための *Caltech 101 (Fei-Fei et al., 2004)* や、画像分類と物体検出のための *PASCAL VOC (Everingham et al., 2012)* などのデータセットでは、すでにいくつかの指標が確立されています。これらの手法を大規模環境に適応させるために、私たちは3つの重要な課題に対処しなければなりませんでした。

1. 画像分類と単一物体の定位では、データセットの規模が大きいため、各画像には1つの物体カテゴリしかラベル付けできませんでした。このため、評価の際に曖昧さが生じる可能性がありました（セクション4.1参照）。
2. オブジェクトのクラスタを含む一部の画像では、オブジェクトインスタンスの定位を評価することが困難である（4.2 節参照）。
3. 画像中のピクセル数が少ないオブジェクトインスタンスの定位評価は困難である（4.3 節参照）。

このセクションでは、3 つの ILSVRC タスクのそれぞれについて、標準化された評価基準を説明します。また、大規模な評価を行う上でのこれらの課題やその他の細かい課題についても詳しく説明しています。付録 E には、投稿プロトコルとその他の競技会運営の詳細を記述します。

4.1. Image classification

ILSVRC の分類タスクの規模（1000 のカテゴリと 100 万枚以上の画像）は、すべての画像に含まれるすべてのオブジェクトのすべてのインスタンスにラベルを付けることを非常に高価にしています。そのため、このデータセットでは、各画像に1つのオブジェクトカテゴリのみがラベル付けされています。これは評価に曖昧さをもたらします。例えば，ある画像が「いちご」とラベル付けされていても，いちごとりんごの両方が含まれているかもしれません． そうすると、アルゴリズムは2つのオブジェクトのうちどちらに名前を付けるべきかわかりません。画像の分類タスクでは、アルゴリズムが画像中の複数（最大5個）のオブジェクトを識別することを許可し、オブジェクトの1つが実際に基底真実ラベルに対応している限り、ペナルティを受けないようにしました。図7(上段)にいくつかの例を示します。

5. Methods

ILSVRCデータセットとコンペティションにより、大規模な画像の認識と検索における重要なアルゴリズムの進歩が可能になりました。

5.1. Challenge entries

このセクションでは、毎年ILSVRCに参加した特に革新的で成功した方法を時系列で紹介しています。表 5, 6, 7 に参加した全チームのリストを示す。2012年には、大規模なコンボリューショナル・ニューラルネットワークの開発で転機が訪れています。

ILSVRC 2010

初年度の課題は画像分類(Image classification)のみでした．受賞したNECチーム(Linら, 2011)は、SIFT(Lowe, 2004)とLBP(Ahonenら, 2006)の特徴量と2つの非線形符号化表現(Zhouら, 2010; Wangetら., 2010)と確率的SVMを使用した。XRCEチーム（Perronninら， 2010）は、実証済みのフィッシャーベクトル表現（Perronnin and Dance,2007）を使用し、PCAによる次元削減とデータ圧縮を行った後、線形SVMを使用した。Fisher ベクトルベースの手法は、5年間の研究期間を経て進化し、2010年から2014年までのすべてのILSVRCで強力な性能を発揮し続けてきました。

ILSVRC 2011

2011年に優勝したのは、2010年の準優勝チームXRCEで、高次元画像シグネチャ（Perronninら、2010年）に積量子化を用いた圧縮（Sanchezand Perronnin、2011年）と一対他の線形SVMを適用しています。単一物体定位競争(Single-object localization)は、その年に初めて開催され、2つの勇敢なエントリーがあった。優勝したのはUvAチームで、クラスに依存しないオブジェクト仮説領域を生成するための選択的探索アプローチを用いた（van de Sandeら、2011b）、その後、いくつかのカラーSIFT特徴の高密度サンプリングとベクトル量子化（van de Sandeら、2010）、空間ピラミッドマッチングによるプーリング（Lazebnikら、2006）、GPU上で訓練されたヒストグラム交差カーネルSVM（Majiand Malik、2009）による分類（van de Sandeら、2011a）が行われました。

ILSVRC 2012

大規模物体認識のターニングポイントとなったのが、大規模ディープニューラルネットの登場です。2012年に、画像分類(Image classification)と単一オブジェクトの定位(Single-object localization) の両方のタスクで文句なしの勝者となったのは、SuperVisionチームでした。彼らは、効率的なGPU実装と新しい隠れユニットのドロップアウトトリックを用いて、6000万個のパラメータを持つ大規模なディープ畳み込みニューラルネットワークをRGB値で訓練しました(Krizhevskyら.,2012; Hintonら.,2012)。画像分類の第2位はISIチームで、フィッシュベクトル(Sanchez and Perronnin, 2011)とグラフィカルガウスベクトルのストリームライン版(Harada and Kuniyoshi, 2012)を使用し、パッシブアグレッシブ(PA)アルゴリズム(Crammerら,2006)を使用した線形分類器を使用しました。

単一物体定位の第2位はVGGであり、高密度SIFT特徴量とカラー統計量（Lowe, 2004）、フィッシャーベクトル表現（Sanchezand Perronnin, 2011）、線形SVM分類器、さらに(Arandjelovic and Zisserman, 2012; Sanchezら, 2012)の知見を加えた画像分類システムである。

ISIとVGGは物体の定位に(Felzenszwalbら, 2010)を使用し、SuperVisionはバウンディングボックスの位置を予測するために訓練された回帰モデルを使用した。検出モデルが弱いにもかかわらず、SuperVisionは物体定位タスクに勝利しました。単一物体定位タスクにおけるSuperVisionとVGGの詳細な分析と比較については、(Russakovskyら, 2013)を参照してください。SuperVisionモデルの成功の影響は、ILSVRC2013とILSVRC2014ではっきりと見ることができます。

ILSVRC 2013

今回のILSVRC2013には24チームが参加しましたが、3年前の大会では21チームが参加しました。2012年の深層学習法の成功に続き、2013年の応募作品の大半は深層畳み込みニューラルネットワークを用いたものでした。画像分類タスクの勝者は、いくつかの大規模な深層畳み込みニューラルネットワークを平均化した *Clarifai* であった。ネットワークアーキテクチャは、可視化手法(Zeiler and Fergus, 2013)を用いて選択され、ドロップアウト技術(Krizhevsky et al., 2012)を用いてGPU上で訓練された(Zeiler et al., 2011)。

受賞した「Over Feat」は、マルチスケールのスライディングウィンドウアプローチを用いた分類、定位、検出のための畳み込みネットワークを使用する統合的なフレームワークに基づいています（Sermanetら, 2013）。 このチームは、3つのタスクすべてに取り組んだ唯一のチームでした。

物体検出タスクの勝者は、UvAチームであり、多値空間ピラミッドを用いてプールされた高密度サンプリングされた色記述子(van de Sandeら, 2010)を効率的に符号化する新しい方法(van de Sandeら, 2014)を選択的探索フレームワーク(Uijlingsら, 2013)で利用していた。検出結果は、フル画像畳み込みネットワーク分類器を用いて再スコア化した。

ILSVRC 2014

2014年は最も多くの応募があり、2013年のわずか24チームに対し、36チームが123の応募を行いました。2013年と同様に、ほぼすべてのチームが畳み込みニューラルネットワークスをベースにして応募しました。画像分類の誤差は ILSVRC2013 からほぼ半減し、物体検出の平均精度は ILSVRC2013 と比較して約 2 倍になりました。詳細は 6.1 節を参照してください。

2014年には、外部データの利用が認められたため、画像分類、単一物体定位、物体検出の各タスクにおいて、提供データトラックと外部データトラックの6つのトラックが用意されました。

提供されたデータを用いた画像分類で優勝したのはGoogLeNetで、ヘブビアン原理から得られた直感とマルチスケールの考え方を組み合わせた改良されたコンボリューショナル・ニューラル・ネットワーク・アーキテクチャを探索しました。 次元削減層を追加することで、計算オーバーヘッドを大幅に発生させることなく、ネットワークの深さと幅の両方を大幅に増加させることができました。 外部データトラックを用いた画像分類では、CASIAWSが分類ラベルのみからの弱教師付き物体定位を用いて画像分類を改善し、勝利した。PASCAL VOC 2012データ上で事前に訓練されたMCG領域提案（Arbel ́aezら、2014）を用いて領域提案を抽出し、畳み込みネットワークを用いて領域を表現し、複数インスタンス学習戦略を用いて弱教師付き物体検出器を学習して画像を表現した。

データトラックを使用した単一オブジェクトの定位では、VGGが優勝しました。このチームは，Caffe (Jia, 2013)の実装をベースに，線形単位非線形性を整流した最大19の重み層を持つ3つの異なるアーキテクチャを使用して，畳み込みニューラルネットワークの深さが精度に与える影響を調査したものであった。定位には，OverFeat (Sermanet et al., 2013)と同様のクラスごとのバウンディングボックス回帰を使用した．外部データトラックを用いた単一物体の定位では，Adobe は 2000 個の ImageNet クラスを追加して，分類と定位の両方のための統合畳み込みニューラルネットワークフレームワークで分類器を訓練し，バウンディングボックス回帰を使用した．テスト時には、k-meansを使用してバウンディングボックスクラスタを見つけ、分類スコアに応じてクラスタをランク付けしました。

また、提供されたデータトラックを用いた物体検出では、優勝したNUSチームはRCNNフレームワーク(Girshickら, 2013)に、(Howard,2014)の改良を加えたNetwork-in-network法(Lin et al., 2014a)を用いました。(Chenら, 2014)に倣ってグローバルコンテキスト情報を取り入れた。外部データトラックを用いた物体検出では、GoogLeNet(提供データを用いた画像分類でも優勝)が優勝しました。同一チームが画像分類と物体検出の両方に成功したことは、シーン情報に基づいて画像を分類するだけでなく、複数の物体を正確に定位させることができることを示しており、注目に値します。このトラックに参加しているほとんどのチームと同様に、GoogLeNetも画像分類データセットを追加のトレーニングデータとして使用しています。

5.2. Large scale paradigm shift

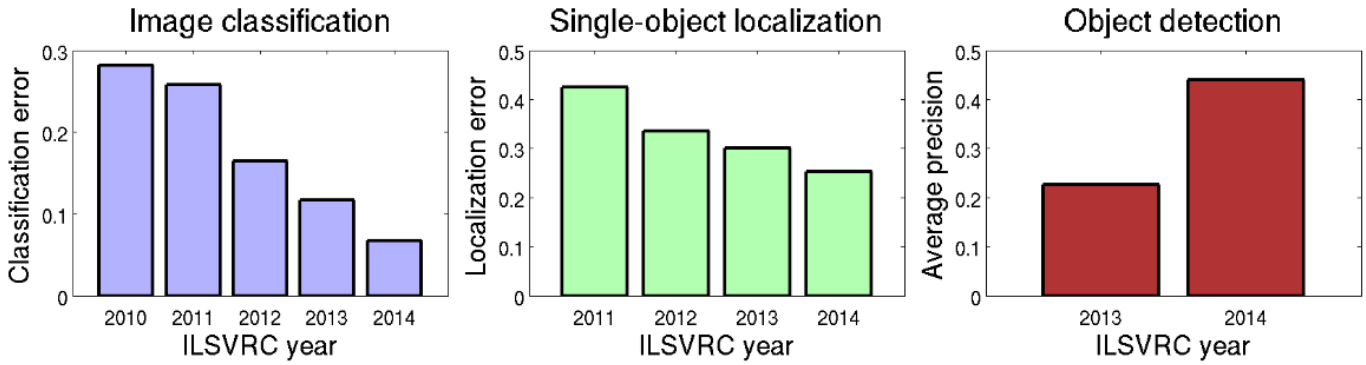
ILSVRCは過去5年間に渡り、コンピュータビジョンの大きなパラダイムシフトの道を切り開いてきました。

カテゴリカルな物体認識の分野は、大規模な設定で劇的な進化を遂げてきた。セクション5.1では、コード化されたSIFT特徴量から始まり、画像分類、単一物体の定位、物体検出の3つのタスクのすべてにおいて、大規模な畳み込みニューラルネットワークが支配するように進化してきた様子を記録しています。膨大な量のトレーニングデータが利用可能になったことで、抽出された特徴と識別的分類器の多段階のハンドチューニングパイプラインを作成する必要なく、画像データから直接ニューラルネットワークを学習することが可能になりました。2012年には、画像分類と単一物体の定位タスクでSuperVisionチームが勝利し、大きなブレークスルーとなりました(Krishevskyet al., 2012)。そして2014年までには、上位の出場者全員が畳み込みニューラルネットワークに大きく依存していました。

また，コンピュータビジョンの分野全体では，ここ数年，大規模認識に注目が集まっています． 2013年のトップビジョン会議では、大規模認識手法が最優秀論文賞を受賞しました。CVPR 2013では、"Fast, Accurate Detection of 100,000Object Class on a Single Machine" (Deanら, 2013)、ICCV 2013では、"From Large Scale Image Cate-gorization to Entry-Level Categories" (Ordonezら,2013)が受賞しました。さらに，ECCV2012で最優秀論文賞を受賞した(Kuettelら, 2012)の大規模弱教師付きローカリゼーションや，(Fromeら, 2013)のような大規模ゼロショット学習など，影響力のある研究がいくつか登場している．

6.1. Improvements over the years

ILSVRC2010 から ILSVRC2014 までの間に、最新の精度が大幅に向上しており、過去 5 年間で大規模物体認識が飛躍的に進歩したことを示しています。図 9 にタスク別、年度別の ILSVRC 入賞作品の性能を示す。年を追うごとに改善されていることがわかる。本節では、この改善点を定量化して分析する。



ILSVRC2010-2014 の入賞作品の 3 つのタスクごとの性能（応募作品および数値結果の詳細は 5.1 節に記載）。物体分類と単一物体定位では、毎年着実に誤差が減少しており、物体検出では平均精度が1.9倍向上しています。これらの比較を行う上で考慮すべき点が2つあります。(1) ISLVRCでは、2010年と2011年、2011年と2012年の間に、対象物のカテゴリが変化している。しかし、データの大規模化（1000 個の物体カテゴリ、120 万枚のトレーニング画像）は変わっていないため、結果を比較することが可能である。ここで示した画像分類と単一物体の定位のエントリは、提供されたトレーニングデータのみを使用しています。 (2) 2013 年から 2014 年にかけて、物体検出訓練データのサイズが大幅に増加している（3.3 節）。6.1節では、訓練データの増加とアルゴリズムの改良の相対的な効果について考察する。

6.1.1. 画像分類と単一物体の定位の長年にわたる改善

挑戦を開始してから、画像分類誤差が4.2倍（28.2%→6.7%）、単一物体の定位誤差が1.7倍（42.5%→25.3%）減少しています。 一貫性を持たせるために、ここでは提供されたトレーニングデータを使用したチームのみを考慮しています。正確なオブジェクトのカテゴリが変わっても（3.1.1節）、データの規模は変わらず（表2）、年をまたいで比較可能な結果が得られた。2012年以降はデータセットに変化はなく、過去3年間で画像分類誤差が2.4倍（16.4%→6.7%）、単一物体の定位誤差が1.3倍（33.5%→25.3%）減少しています。

6.1.2. 長年にわたる物体検出の改善

平均平均精度（mAP）で測定した物体検出精度は、ILSVRC2013 では 22.6%だったものが、ILSVRC2014 では 43.9%と、本タスクの導入以来 1.9 倍に向上しています。しかし、これらの結果は2つの理由から直接比較することはできません。第一に、物体検出訓練データのサイズが 2013 年から 2014 年にかけて大幅に増加していることである（第 3.3 節）。 第二に、43.9%のmAPは画像分類と単一物体定位訓練データを追加して得られた結果である。ここでは、訓練セットサイズの増加とアルゴリズムの改善の相対的な効果を理解することを試みる。すべてのモデルは、同じILSVRC2013-2014の物体検出テストセットで評価されています。