

Wissenschaftliches Schreiben

Teil 2 - Tipps für die Praxis

25. November 2019 (WS19/20)



Jonas Jäger¹

¹ Doktorand am Fachbereich Elektrotechnik & Informationstechnik

Overview



1. Allgemeines

2. Wichtige Elemente

3. 10 Tipps

Was mir hilft



- Speedreading
 - Buchempfehlung: Speed Reading, Buzan [2013]
- 10 Finger schreiben
 - Tipp 10
- L^AT_EX+ Bibtex
- Literaturverwaltung (JabRef)

Definition Gliederungsebenen



Titel

1. Abschnitt

1.1 Unterabschnitt

1.1.1 Absatz: Besteht aus mehreren Sätzen

1.1.2 ...

1.2 Unterabschnitt

1.2.1 Absatz

1.2.2 ...

2. Abschnitt

2.1 Absatz

2.2 ...



Beispiel: Gliederungsebenen

with a deep network and training a high-capacity model with only a small quantity of annotated detection data.

Unlike image classification, detection requires localizing (likely many) objects within an image. One approach frames localization as a regression problem. However, work from Szegedy *et al.* [33], concurrent with our own, indicates that this strategy may not fare well in practice (they report a mAP of 30.5% on VOC 2007 compared to the 58.5% achieved by our method). An alternative is to build a sliding-window detector. CNNs have been used in this way for at least two decades, typically on constrained object categories, such as faces [29, 35] and pedestrians [31]. In order to maintain high spatial resolution, these CNNs typically only have two convolutional and pooling layers. We also considered adopting a sliding-window approach. However, units high up in our network, which have five convolutional layers, have very large receptive fields (195×195 pixels) and strides (32×32 pixels) in the input image, which makes precise localization within the sliding-window paradigm an open technical challenge.

Instead, we solve the CNN localization problem by operating within the “recognition using regions” paradigm [19], which has been successful for both object detection [34] and semantic segmentation [5]. At test time, our method generates around 2000 category-independent region proposals for the input image, extracts a fixed-length feature vector from each proposal using a CNN, and then classifies each region with category-specific linear SVMs. We use a simple technique (affine image warping) to compute a fixed-size CNN input from each region proposal, regardless of the region’s shape. Figure 1 presents an overview of our method and highlights some of our results. Since our system combines region proposals with CNNs, we dub the method R-CNN: Regions with CNN features.

A second challenge faced in detection is that labeled data is scarce and the amount currently available is insufficient for training a large CNN. The conventional solution to this problem is to use *unsupervised* pre-training, followed by supervised fine-tuning (e.g., [31]). The second principle contribution of this paper is to show that *supervised* pre-training on a large auxiliary dataset (ILSVRC), followed by domain-specific fine-tuning on a small dataset (PASCAL), is an effective paradigm for learning high-capacity CNNs when data is scarce. In our experiments, fine-tuning for detection improves mAP performance by 8 percentage points. After fine-tuning, our system achieves a mAP of 54% on VOC 2010 compared to 33% for the highly-tuned, HOG-based deformable part model (DPM) [15, 18]. We also point readers to contemporaneous work by Donahue *et al.* [11], who show that Krizhevsky’s CNN can be used (without fine-tuning) as a blackbox feature extractor, yielding excellent performance on several recognition tasks including scene classification, fine-grained sub-categorization, and domain

adaptation.

Our system is also quite efficient. The only class-specific computations are a reasonably small matrix-vector product and greedy non-maximum suppression. This computational property follows from features that are shared across all categories and that are also two orders of magnitude lower-dimensional than previously used region features (cf. [34]).

Understanding the failure modes of our approach is also critical for improving it, and so we report results from the detection analysis tool of Hoiem *et al.* [21]. As an immediate consequence of this analysis, we demonstrate that a simple bounding box regression method significantly reduces mislocalizations, which are the dominant error mode.

Before developing technical details, we note that because R-CNN operates on regions it is natural to extend it to the task of semantic segmentation. With minor modifications, we also achieve competitive results on the PASCAL VOC segmentation task, with an average segmentation accuracy of 47.9% on the VOC 2011 test set.

2. Object detection with R-CNN

Our object detection system consists of three modules. The first generates category-independent region proposals. These proposals define the set of candidate detections available to our detector. The second module is a large convolutional neural network that extracts a fixed-length feature vector from each region. The third module is a set of class-specific linear SVMs. In this section, we present our design decisions for each module, describe their test-time usage, detail how their parameters are learned, and show results on PASCAL VOC 2010–12.

2.1. Module design

Region proposals. A variety of recent papers offer methods for generating category-independent region proposals. Examples include: objectness [1], selective search [34], category-independent object proposals [12], constrained parametric min-cuts (CPMC) [5], multi-scale combinatorial grouping [3], and Cireşan *et al.* [6], who detect mitotic cells by applying a CNN to regularly-spaced square crops, which are a special case of region proposals. While R-CNN is agnostic to the particular region proposal method, we use selective search to enable a controlled comparison with prior detection work (e.g., [34, 36]).

Feature extraction. We extract a 4096-dimensional feature vector from each region proposal using the Caffe [22] implementation of the CNN described by Krizhevsky *et al.* [23]. Features are computed by forward propagating a mean-subtracted 227×227 RGB image through five convolutional layers and two fully connected layers. We refer readers to [22, 23] for more network architecture details.

Wie man ein Paper liest



1. Titel
2. Überblick + Abbildungen
3. Kurzfassung + Schlussfolgerung/ Zusammenfassung
4. Einführung
5. Den Rest lesen



soaktuell.ch



salzburg.com

Titel¹



- Interesse des Lesers wecken
- Prägnante Darstellung des Beitrags
- Abgrenzung von anderen Titeln
- Bietet *Schlüsselworte* für Suchmaschinen



muebri.de

¹Lebrun [2007]

Titel: Leitfragen²



- Was für ein Problem wird behandelt?
- Welche Methoden werden zur Lösung des Problems verwendet?
- Welche Daten werden verwendet und was ist die Applikation?
- Beispiel:

Rich feature hierarchies for accurate object detection and semantic segmentation

Beispieldatum aus Girshick et al. [2013]

²Paper Preparation Guidelines - CVG Jena

Kurzfassung³



- Macht den Titel verständlich
 - Schlüsselwörter des Titels aufgreifen
- Beschreibt den wissenschaftlichen Beitrag
- Leser: Hilft bei der Leseentscheidung
- Besteht aus 4 Teilen
 - Problem, Lösung, Ergebnisse, Beitrag

³Lebrun [2007]

Kurzfassung: Leitfragen⁴



1. Problem

- Was ist das Problem?
- Was ist das Thema des Papers?

2. Lösung

- Wie wurde das Problem gelöst?

3. Ergebnisse

- Was sind die Ergebnisse?
- Wie gut wurde das Problem gelöst?

4. Beitrag

- Wie nützen diese Ergebnisse der Wissenschaft oder dem Leser?

⁴Lebrun [2007]

Kurzfassung: Beispiel



Abstract

Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a mAP of 53.3%. Our approach combines two key insights: (1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since we combine region proposals with CNNs, we call our method R-CNN: Regions with CNN features. Source code for the complete system is available at <http://www.cs.berkeley.edu/~rbg/rcnn>.

Kurzfassung aus Girshick et al. [2013]

Einführung⁵



- Schreibstil: Aktiv (persönlich)
- Erzählt eine Geschichte
- Motivation
- Sorgt für Vertrauenswürdigkeit
 - Grenzt das Anwendungsgebiet ein
- Versorgt den Leser mit Kontext

⁵Lebrun [2007]

Einführung: Leitfragen⁶



- Warum jetzt?
- Warum dieses Problem? Warum ist dieses Problem wichtig?
- Warum wurde das Problem auf diesem Weg gelöst?
- Welche Methoden wurden verwendet?
- Wie unterscheidet sich diese Arbeit von Anderen?
- Warum ist der Beitrag für den Leser interessant?
- Was sind mögliche Anwendungen?

⁶Lebrun [2007]

Experimente⁷



- Beweist/ Stützen die Argumentation
- Systematische Evaluation des Beitrags
- Best practice
 - 1. Ein Schlüsselexperiment
 - 2. Ergebnis in einer Grafik (Graph) veranschaulichen
 - 3. Diese Grafik zu Beginn des Abschnitts referenzieren
 - 4. Bildunterschrift der Grafik muss ohne den restlichen Text verständlich sein



⁷Paper Preparation Guidelines - CVG Jena

Experimente: Leitfragen⁸



- Was möchte ich evaluieren?
- Welche Ergebnisse sollen veranschaulicht?
- Was ist der beste Standard für die Bewertung meiner Ergebnisse?
- Sind alle Bildunterschriften verständlich, auch ohne den Text zu lesen?



⁸Paper Preparation Guidelines - CVG Jena

Experimente: Beispiel - quantitative Ergebnisse



VOC 2010 test	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
DPM v5 [18] [†]	49.2	53.8	13.1	15.3	35.5	53.4	49.7	27.0	17.2	28.8	14.7	17.8	46.4	51.2	47.7	10.8	34.2	20.7	43.8	38.3	33.4
UVA [34]	56.2	42.4	15.3	12.6	21.8	49.3	36.8	46.1	12.9	32.1	30.0	36.5	43.5	52.9	32.9	15.3	41.1	31.8	47.0	44.8	35.1
Regionlets [36]	65.0	48.9	25.9	24.6	24.5	56.1	54.5	51.2	17.0	28.9	30.2	35.8	40.2	55.7	43.5	14.3	43.9	32.6	54.0	45.9	39.7
SegDPM [16] [†]	61.4	53.4	25.6	25.2	35.5	51.7	50.6	50.8	19.3	33.8	26.8	40.4	48.3	54.4	47.1	14.8	38.7	35.0	52.8	43.1	40.4
R-CNN	67.1	64.1	46.7	32.0	30.5	56.4	57.2	65.9	27.0	47.3	40.9	66.6	57.8	65.9	53.6	26.7	56.5	38.1	52.8	50.2	50.2
R-CNN BB	71.8	65.8	53.0	36.8	35.9	59.7	60.0	69.9	27.9	50.6	41.4	70.0	62.0	69.0	58.1	29.5	59.4	39.3	61.2	52.4	53.7

Table 1: Detection average precision (%) on VOC 2010 test. R-CNN is most directly comparable to UVA and Regionlets since all methods use selective search region proposals. Bounding box regression (BB) is described in Section 3.4. At publication time, SegDPM was the top-performer on the PASCAL VOC leaderboard. [†]DPM and SegDPM use context rescoring not used by the other methods.

Ergebnistabelle im Detectionsgenauigkeiten aus Girshick et al. [2013]

Experimente: Beispiel - qualitative Ergebnisse

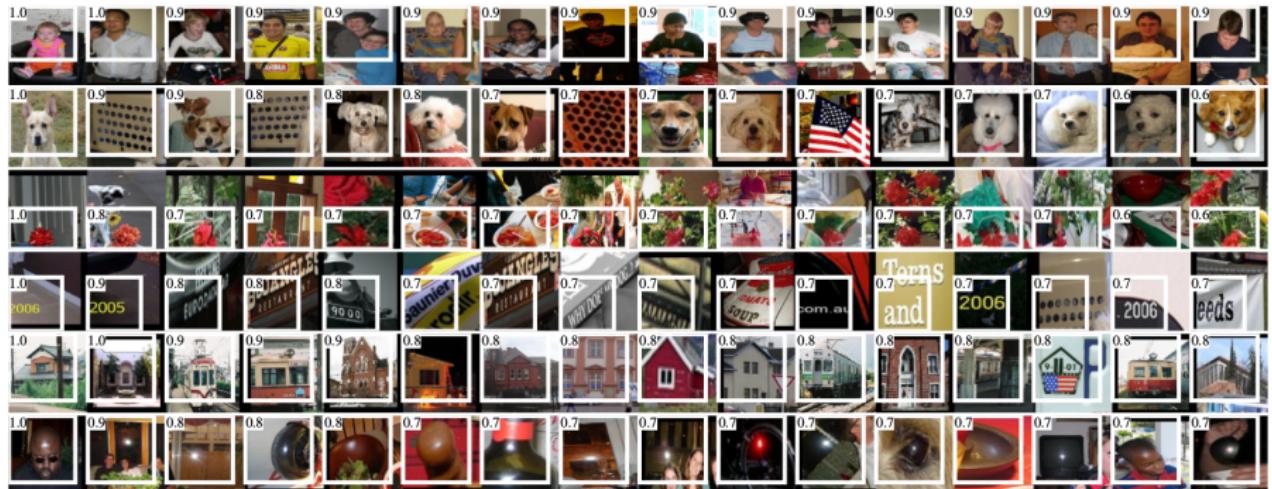


Figure 3: Top regions for six pool₅ units. Receptive fields and activation values are drawn in white. Some units are aligned to concepts, such as people (row 1) or text (4). Other units capture texture and material properties, such as dot arrays (2) and specular reflections (6).

Beispieldetektionen aus Girshick et al. [2013]

Schlussfolgerung⁹



- Zusammenfassung: Hauptaspekte der Arbeit
 - Referenzieren wichtiger Aspekte (Abbildungen, Formeln, ...)
- Benennt den Beitrag in anderen Worten
 - Schwerpunkt: Nutzen für Andere
- Gegenstück zur Kurzfassung und Einführung



carleton.ca

⁹Paper Preparation Guidelines - CVG Jena

Schlussfolgerung: Leitfragen¹⁰



- Was ist das Ergebnis meiner Arbeit?
- Was sind die wichtigsten Fakten, die nochmal erwähnt werden sollen?
- Was sind Schlüsselaspekte (Grafiken, Formeln, ...) die hier referenziert werden sollen, da sie den Hauptanteil der Lösung ausmachen?
- Was könnte potenzielle Kritik sein? Habe ich diese Probleme behandelt? Ist hier weitere Arbeit zu leisten um diese Probleme zu lösen?
- Was sind meine besten Ergebnisse? Gibt es Auswertungen in meinem Paper die diese Ergebnisse beweisen?

¹⁰Paper Preparation Guidelines - CVG Jena

Überschriften¹¹



Drei Prinzipien:

1. Der Beitrag bestimmt die Form
 - Der detaillierteste Abschnitt sollte den Beitrag betreffen
 - Der Abschnitt mit den meisten Unterabschnitten sollte den Beitrag betreffen
2. Schlüsselworte des Titels werden in Überschriften wiederholt
3. Überschriften unterstützen die Geschichte
 - Sind in logischer Reihenfolge

¹¹Lebrun [2007]

Schaubilder



- Beantworten wichtige Fragen
 - Werfen keine Fragen auf!
- Sind auf das wesentliche reduziert
- Bild + Bildunterschrift = Verständlich
 - Müssen ohne den Text verständlich sein!
- Werden im Text referenziert und beschrieben

1. Verwende einfache Sprache¹²



- Kurze Sätze
- Logische Struktur
- Fachjargon auf ein Minimum begrenzen

KISS – Keep It Simple, Stupid!

„Wenn Du es einem
Sechsjährigen nicht
erklären kannst, hast du
es selbst noch nicht gut
genug verstanden.“
Albert Einstein



Einfach schön, Zen-Prinzip – Dirk Hannemann, Berlin

79

Einfach schön - Dirk Hannemann, Berlin

¹²Nowozin, Ten Tips for Writing CS Papers - Part 1

2. Benenne den Beitrag¹³



- Beitrag = Problem + Lösung
- Mögliche Beiträge
 - 1. Sie haben eine Erklärung für etwas das bereits existiert
 - 2. Sie haben etwas verbessert
 - 3. Sie haben etwas entwickelt, das es vorher nicht gab
 - 4. Sie wenden einen vorhandenen Lösungs...neues Problem an



kaizeninstituteindia.wordpress.com

¹³Nowozin, Ten Tips for Writing CS Papers - Part 1

3. Betrachte alles als eine Facette des Beitrags¹⁴



1. Beitrag klar darstellen
2. Jeder Abschnitt unterstützt
 - *Einführung*: Motiviert den Beitrag
 - *Stand der Technik*: Grenzt Ihren Beitrag zu anderen Arbeiten ab
 - *Methode*: Beschreibung des Beitrags
 - *Experimente*: Bestätigt, dass Ihr Beitrag funktioniert



¹⁴Nowozin, Ten Tips for Writing CS Papers - Part 1

4. Verwende ein Seite-1-Schaubild¹⁵



- Visuelle Kurzfassung
- Eigenschaften
 - Verständlich in 20 Sekunden
 - Selbsterklärend (Bild + Bildunterschrift)

¹⁵Nowozin, Ten Tips for Writing CS Papers - Part 1

Seite-1-Schaubild: Beispiel



R-CNN: *Regions with CNN features*

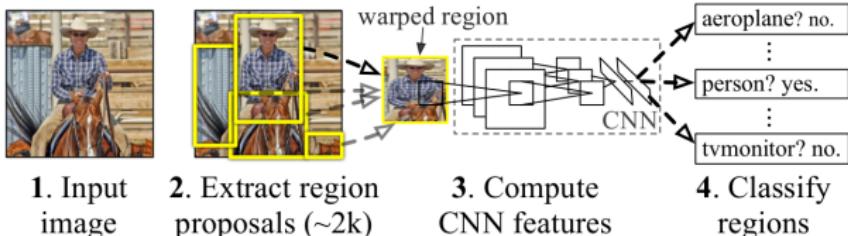


Figure 1: Object detection system overview. Our system (1) takes an input image, (2) extracts around 2000 bottom-up region proposals, (3) computes features for each proposal using a large convolutional neural network (CNN), and then (4) classifies each region using class-specific linear SVMs. R-CNN achieves a mean average precision (mAP) of **53.7% on PASCAL VOC 2010**. For comparison, [34] reports 35.1% mAP using the same region proposals, but with a spatial pyramid and bag-of-visual-words approach. The popular deformable part models perform at 33.4%.

Seite-1-Schaubild aus Girshick et al. [2013]

5. Vermeide das Passiv¹⁶



Die meisten Menschen empfinden das Aktiv als

- leichter zu lesen
- ansprechender/ fesselnder

Weitere Informationen: [Passive Voice in Scientific Writing](#)



<http://www.elternwissen.com>

¹⁶Nowozin, Ten Tips for Writing CS Papers - Part 1

6. Ideale Struktur eines Absatzes¹⁷



- 1. Ein Absatz = Eine Idee**
2. Beginn und Ende des Absatzes verbinden
 - mit dem Kontext
 - mit den einschließenden Absätzen



experto.de

¹⁷Nowozin, Ten Tips for Writing CS Papers - Part 2

6. Ideale Struktur eines Absatzes¹⁸: Beispiel



Our object detection system consists of three modules. The first generates category-independent region proposals. These proposals define the set of candidate detections available to our detector. The second module is a large convolutional neural network that extracts a fixed-length feature vector from each region. The third module is a set of class-specific linear SVMs. In this section, we present our design decisions for each module, describe their test-time usage, detail how their parameters are learned, and show results on PASCAL VOC 2010-12.

Absatz aus Girshick et al. [2013]

¹⁸Nowozin, Ten Tips for Writing CS Papers - Part 2

6. Ideale Struktur eines Absatzes¹⁹



Herangehensweise (pro Abschnitt)

1. Liste von Stichpunkten/ Ideen erstellen
2. Abhängigkeiten zwischen den einzelnen Ideen erkennen
3. Stichpunkte in logische Reihenfolge bringen
4. Einen Absatz pro Idee schreiben

¹⁹Nowozin, Ten Tips for Writing CS Papers - Part 2

7. Vermeide mehrdeutige Verweise/ Demonstrativpronomen²⁰



- Dieser, Dieses , Jener, Jenes, Der, Die, Das,...



bjoernschnabel.wordpress.com

²⁰Nowozin, Ten Tips for Writing CS Papers - Part 2

8. Verwende Ankündigungen²¹



- Ankündigungen
 - Zu Beginn eines Abschnitts
 - Was erwartet den Leser?
- Aufgabe aus Lesersicht
 - Erleichterte Navigation
 - Überspringen
 - Wiedereinstieg

²¹Nowozin, Ten Tips for Writing CS Papers - Part 2

9. Mehrere Autoren²²



- Inhalt
 - 1 Schreiber für den Entwurf (erste Version des Papers)
 - Diesen Entwurf zusammen verfeinern
- Technologie
 1. \LaTeX + Git (Bitbucket, GitLab)
 - Tipp: Eine Zeile pro Satz im Tex-Dokument
 2. PDF + Kommentarfunktion
 3. Word + „Änderungen Nachverfolgen“
 4. Office 365 Cloud
 5. Share \LaTeX

²²Nowozin, Ten Tips for Writing CS Papers - Part 2

10. Autoren-Reihenfolge²³



- Reihenfolge
 - Student1, Student2, Betreuer, Professor
- Wer wird Hauptautor?
- Wer sind die Autoren?
- Beispiel:

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik
UC Berkeley

{rgb, jdonahue, trevor, malik}@eecs.berkeley.edu

Beispiel Autoren-Reihenfolge Girshick et al. [2013]

²³Nowozin, Ten Tips for Writing CS Papers - Part 2

Danke für Ihre Aufmerksamkeit



„Good writing is bad writing that was rewritten.“ - Marc H. Raibert

References I



- T. Buzan. *Speed reading : schneller lesen – mehr verstehen – besser behalten.* mvg-Verl., München, 1., aktualisierte und erw. aufl. edition, 2013. ISBN 3868824383. URL
http://deposit.d-nb.de/cgi-bin/dokserv?id=4228259&prov=M&dok_var=1&dok_ext=htm.
- R. B. Girshick, J. Donahue, T. Darrell, and J. Maliks. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.
- J.-L. Lebrun. *Scientific Writing: A Reader and Writer's guide.* World Scientific, 2007.