Title: Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods

Article preview

Pattern Recognition
Volume 71, November 2017, Pages 132-143
Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods
Author links open overlay panelMassimiliano Patacchiola, Angelo Cangelosi

https://doi.org/10.1016/j.patcog.2017.06.009Get rights and content
Highlights
 * ?
A convolutional neural network approach for head pose estimation is proposed.
 * ?
The performance of different network architectures has been measured.
 * ?
The use of adaptive gradient methods leads to the state-of-the-art in wild datasets.
 * ?
We release a library based on our work which is available under open source licence.
Abstract
Head pose estimation is an old problem that is recently receiving new attention because of possible applications in human-robot interaction, augmented reality and driving assistance. However, most of the existing work has been tested in controlled environments and is not robust enough for real-world applications. In order to handle these limitations we propose an approach based on Convolutional Neural Networks (CNNs) supplemented with the most recent techniques adopted from the deep learning community. We evaluate

the performance of four architectures on recently released in-the-wild datasets. Moreover, we investigate the use of dropout and adaptive gradient methods giving a contribution to their ongoing validation. The results show that joining CNNs and adaptive gradient methods leads to the state-of-the-art in unconstrained head pose estimation.

Introduction

In the last few years major advancements in robotics, augmented reality and driving assistance have highlighted the need for robust methods to estimate the head pose in real-world scenarios. For instance, robots are gradually leaving factories and becoming part of our lives as companions and as assistants. It has been shown that in human-robot interaction a coarse pose estimation of the head is a fundamental prerequisite for building trust with users during joint-attention tasks [1]. In the context of autonomous cars a driving assistance system could take advantage of head pose estimation for decelerating the car when pedestrians do not notice the presence of the vehicle [2]. Moreover a similar system can be installed inside the vehicle and used to monitor the driver?s awareness. The need of a robust head pose estimation is not limited to these domains. There have been significant applications in surveillance and anomaly detection, human-computer interaction and crowd behavioural dynamics analysis [3]. All of these unconstrained scenarios need an estimator which is resistant to variable environmental conditions, and which can evaluate the focus of attention in absence of more accurate information such as the gaze. Here it is necessary to specify what we consider as a wild environment. We define as taken in a wild environment those face images exhibiting a large variety in appearance (pose, expression, ethnicity, age, gender, etc.), environmental conditions (artificial light, shadows, etc.), and containing relevant occlusions (sunglasses, masks, scarves, etc.). We will show how Convolutional Neural Networks (CNNs) can be considered one of the best algorithms for robust head pose estimation in a wild environment.

We can summarise the main contribution of our work in three points:

* 1.
As far as we know this is the first work that has deeply investigated the use of CNNs in head pose estimation. Our main contribution is a rigorous evaluation of multiple CNN models and factors. The results are compared with other algorithms, and show how an approach based on CNNs, dropout and adaptive gradient methods represents the state of the art in head pose estimation.
* 2.
Deep learning is a rapidly growing field, which is bringing new techniques that can significantly improve the performance of CNNs. Because these techniques have been released in the last few years, there is still a validation process for establishing their cross-domain usefulness. We explored the role of adaptive gradient methods and we gave a valuable contribution to their ongoing validation.
* 3.
The results obtained in this work have been used to implement a Python library called Deepgaze. The library includes pre-trained CNNs based on Tensorflow [4] which can run in real-time on GPUs and mobile devices. Deepgaze is released under an open-source license and is available for both academic and commercial purposes. The software is available on the author?s repository.1

Section snippets

Related work

The head pose estimation problem has been investigated from different points of view and with different techniques. Devices such as laser pointers, camera

arrays, stereo-cameras, magnetic and inertial sensors, have been used to get a stable estimation in controlled situations [5]. More recently some good results have been obtained with commercial depth cameras [6]. However the use of these devices is not always feasible due to space constraints and to technical problems when operating outdoors.

Convolutional Neural Networks

In recent years deep convolutional networks have showed their strength in numerous pattern recognition contests. Some remarkable achievements have been recently obtained in object detection [22], facial expression recognition [23] and scene classification [24]. This technology is increasingly used in commercial applications such as content filtering in social networks, recommendation systems in e-commerce websites or image classifiers in web-search engines. The deep learning revolution has been

Experiments

In this section we report the results obtained using CNNs, dropout and adaptive gradient methods on three public datasets: the Prima head-pose dataset [32], the Annotated Facial Landmarks in the Wild (AFLW) dataset [33], and the Annotated Face in the Wild (AFW) dataset [34]. The former is a well-known dataset which has been around for more than ten years, and it is considered a classic benchmark for head pose algorithms. The second is a recently released in-the-wild dataset, and it has the

Conclusions

In this article we introduced the use of dropout and adaptive gradient methods for head pose estimation with CNNs. Our approach is significantly different from previous research [17], [18], [20], [21], and show how using the most recent deep learning techniques leads to the state-of-the-art in constrained and unconstrained datasets. Our method should be considered as part of a broader system, in particular it can be used in conjunction with a face detector. We implemented the system in Python

Acknowledgment

This material is based upon work supported by the Air Force Office of Scientific Research grant number: A9550-15-1-0025, Air Force Materiel Command, USAF under Award No. FA9550-15-1-0025.

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

Massimiliano Patacchiola attended his studies at La Sapienza University (Rome). After his internship at the Laboratory of Artificial Life and Robotics (Rome), in 2012 he started working as robotics engineer at Eurolink Systems group (Rome), where he spent more than two years creating algorithms and designing systems for the control of UGV (Unmanned Ground Vehicle) and UAV (Unmanned Aerial Vehicle). In 2015 he started a PhD program in robotics and computational modelling at Plymouth University.

Recommended articles

References (47)

* C. Wang _et al._
Robust head pose estimation via supervised manifold learning
Neural Netw.
(2014)
* J. Wu _et al._
A two-stage head pose estimation framework and evaluation
Pattern Recognit.
(2008)
* J. Sherrah _et al._
Face distributions in similarity space under varying head pose

Image Vision Comput.
(2001)
* Q. Liu _et al._
Robust facial landmark tracking via cascade regression
Pattern Recognit.
(2017)
* X. Jin _et al._
Face alignment by robust discriminative hough voting
Pattern Recognit.
(2016)
* K. Nogueira, O.A. Penatti, J.A.d. Santos, Towards better exploiting convolutional neural networks for remote sensing...
* D.E. Rumelhart _et al._
Learning internal representation by error propagation.
Nature
(1986)
* W. Yang _et al._
Dropsample: a new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten chinese character recognition
Pattern Recognit.
(2016)
* E. Ricci _et al._
Learning large margin likelihoods for realtime head pose tracking
16th IEEE International Conference on Image Processing (ICIP)
(2009)
* D. Zanatto _et al._
Priming anthropomorphism: can the credibility of humanlike robots be transferred to non-humanlike robots?
Proceedings of the Eleventh Annual ACM/IEEE International Conference on Human-Robot Interaction.
(2016)
* D. Geronimo _et al._
Survey of pedestrian detection for advanced driver assistance systems
IEEE Trans. Pattern Anal. Mach. Intell.
(2010)
* R.H. Baxter _et al._
An adaptive motion model for person tracking with instantaneous head-pose features
IEEE Signal Process. Lett.
(2015)
* M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat,...
* E. Murphy-Chutorian _et al._
Head pose estimation in computer vision: a survey
IEEE Trans. Pattern Anal. Mach. Intell.
(2009)
* G. Fanelli _et al._
Real time head pose estimation from consumer depth cameras
Pattern Recognit.
(2011)
* D. Lia _et al._
A central profile-based 3D face pose estimation
Pattern Recognit.

(2014)
* P. Martins _et al._
Monocular head pose estimation
International Conference Image Analysis and Recognition
(2008)
* J. Tu _et al._
Evaluation of head pose estimation for studio data
Lect. Notes Comput. Sci.
(2007)
* K. Sundararajan _et al._
Head pose estimation in the wild using approximate view manifolds
The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Boston, Massachusetts
(2015)
* R. Stiefelhagen
Estimating head pose with neural networks - results on the pointing04 icpr workshop evaluation data
Proceedings of Pointing, ICPR, International Workshop on Visual Observation of Deictic Gestures, Cambridge, UK
(2004)
* M. Osadchy _et al._
Synergistic face detection and pose estimation with energy-based models
J. Mach. Learn. Res.
(2007)
* B. Ahn _et al._
Real-time head orientation from a monocular camera using deep neural network
12th Asian Conference on Computer Vision
(2014)
* G. Fanelli _et al._
Random forests for real time 3D face analysis
Int. J. Comput. Vision
(2012)
View more references
Cited by (191)
* ### Anisotropic angle distribution learning for head pose estimation and attention understanding in human-computer interaction
2021, Neurocomputing
Show abstract
Head pose estimation is an important way to understand human attention in the human-computer interaction. In this paper, we propose a novel anisotropic angle distribution learning (AADL) network for head pose estimation task. Firstly, two key findings are revealed as following: 1) Head pose image variations are different at the yaw and pitch directions with the same pose angle increasing on a fixed central pose; 2) With the fixed angle interval increasing, the image variations increase firstly and then decrease in yaw angle direction. Then, the _maximum a posterior_ technology is employed to construct the head pose estimation network, which includes three parts, such as convolutional layer, covariance pooling layer and output layer. In the output layer, the labels are constructed as the anisotropic angle distributions on the basis of two key findings. And the anisotropic angle distributions are fitted by the 2D Gaussian-like distributions (groundtruth labels). Furthermore, the Kullback-Leibler divergence is selected to measure the predication label and the groundtruth one. The features of head pose

images are perceived at the AADL-based convolutional neural network in an end-to-end manner. Experimental results demonstrate that the developed AADL-based labels have several advantages, such as robustness for head pose image missing, insensitivity for the motion blur. Moreover, the proposed method has achieved good performance compared to several state-of-the-art methods on the Pointing?04 and CAS_PEAL_R1 databases.

* ### Recent advances in convolutional neural networks
2018, Pattern Recognition
Show abstract
In the last few years, deep learning has led to very good performance on a variety of problems, such as visual recognition, speech recognition and natural language processing. Among different types of deep neural networks, convolutional neural networks have been most extensively studied. Leveraging on the rapid growth in the amount of the annotated data and the great improvements in the strengths of graphics processor units, the research on convolutional neural networks has been emerged swiftly and achieved state-of-the-art results on various tasks. In this paper, we provide a broad survey of the recent advances in convolutional neural networks. We detailize the improvements of CNN on different aspects, including layer design, activation function, loss function, regularization, optimization and fast computation. Besides, we also introduce various applications of convolutional neural networks in computer vision, speech and natural language processing.

* ### ARHPE: Asymmetric Relation-Aware Representation Learning for Head Pose Estimation in Industrial Human-Computer Interaction
2022, IEEE Transactions on Industrial Informatics

* ### Quatnet: Quaternion-based head pose estimation with multiregression loss
2019, IEEE Transactions on Multimedia

* ### Facial Landmark Detection: A Literature Survey
2019, International Journal of Computer Vision

* ### Fine-grained head pose estimation without keypoints
2018, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops
View all citing articles on Scopus
Massimiliano Patacchiola attended his studies at La Sapienza University (Rome). After his internship at the Laboratory of Artificial Life and Robotics (Rome), in 2012 he started working as robotics engineer at Eurolink Systems group (Rome), where he spent more than two years creating algorithms and designing systems for the control of UGV (Unmanned Ground Vehicle) and UAV (Unmanned Aerial Vehicle). In 2015 he started a PhD program in robotics and computational modelling at Plymouth University. He is currently designing the social skills of different humanoid robots.
Angelo Cangelosi is professor of Artificial Intelligence and Cognition and the Director of the Centre for Robotics and Neural Systems at Plymouth University (UK). Cangelosi studied psychology and cognitive science at the Universities of Rome La Sapienza and at the University of Genoa, and was visiting scholar at the University of California San Diego and the University of Southampton. Cangelosi?s main research expertise is on language grounding and embodiment in humanoid robots, developmental robotics, human-robot interaction, and on the application of neuromorphic systems for robot learning.
View full text

Recommended articles

* ### The Magnification of Atomic Lines Intensity Originated by laser Breakdown in Ultrasound Field

Physics Procedia, Volume 86, 2017, pp. 147-151
A.V. Bulanov, I.G. Nagorny
* ### Disco spotlight sign: A sign that may be useful for diagnosis of pneumomediastinum
The American Journal of Emergency Medicine, Volume 35, Issue 11, 2017, pp. 1786.e1-1786.e2
Yi-Chin Chen, ?, Tzu-Yao Hung
* ### Characterization of high-intensity plaques on noncontrast T1-weighted magnetic resonance imaging by coronary angioscopy
Journal of Cardiology, Volume 70, Issue 6, 2017, pp. 520-523
Akira Oshita, ?, Hiroshi Matsuoka
* ### Partial orders for zero-sum arrays with applications to network theory
Journal of Informetrics, Volume 11, Issue 1, 2017, pp. 257-274
Yuxian Liu, ?, Leo Egghe
* ### Recommendation-Aware Smartphone Sensing System
Journal of Applied Research and Technology, Volume 12, Issue 6, 2014, pp. 1040-1050
Mu-Yen Chen, ?, Hsien-En Lin
* ### Flocking Analysis and Comparison in Simplex Multi-Agent Systems under Matrix-Weighting Hölder Norms
IFAC-PapersOnLine, Volume 48, Issue 28, 2015, pp. 438-443
J. Jhang, ?, X.B. Lu
Show 3 more articles
Article Metrics
Citations
* Citation Indexes191
Captures
* Readers179
View details
* About ScienceDirect
* Remote access
* Shopping cart
* Advertise
* Contact and support
* Terms and conditions
* Privacy policy
Cookies are used by this site. Cookie Settings

Manage Consent Preferences
Strictly Necessary Cookies
Always active
These cookies are necessary for the website to function and cannot be switched off in our systems. They are usually only set in response to actions made by you which amount to a request for services, such as setting your privacy preferences, logging in or filling in forms. You can set your browser to block or alert you about these cookies, but some parts of the site will not then work. These cookies do not store any personally identifiable information.
Cookie Details List?
Functional Cookies
Functional Cookies
These cookies enable the website to provide enhanced functionality and personalisation. They may be set by us or by third party providers whose services we have added to our pages. If you do not allow these cookies then some or all of these services may not function properly.
Cookie Details List?
Performance Cookies
Performance Cookies
These cookies allow us to count visits and traffic sources so we can measure and improve the performance of our site. They help us to know which pages are the most and least popular and see how visitors move around the site.
Cookie Details List?
Targeting Cookies
Targeting Cookies
These cookies may be set through our site by our advertising partners. They may be used by those companies to build a profile of your interests and show you relevant adverts on other sites. If you do not allow these cookies, you will experience less targeted advertising.
Cookie Details List?
Back Button
Cookie List
Search Icon
Filter Icon
Clear
checkbox label label
Apply Cancel
Consent Leg.Interest
checkbox label label
checkbox label label
checkbox label label
Confirm my choices