International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

DEEP POSE AND HUMAN POSE ESTIMATION VIA NEWRAL NETWORK

[1]T Rajesh,[2] Karamsetty Vandana,[3] Yenumula Vyshnavi, [4]Karalapati Pavani, [5]Sangalaprudhvireddy
[1,2,3,4,5]Assistant Professor, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

We propose a method for human pose estimation based on Deep Neural Networks (DNNs). The pose estimation is formulated as a DNN-based regression problem towards body joints. We present a cascade of such DNN regressors which results in high precision pose estimates. The approach has the advantage of reasoning about pose in a holistic fashion and has a simple but yet powerful formulation which capitalizes on recent advances in Deep Learning. We present a detailed empirical analysis with state-ofart or better performance on four academic benchmarks of diverse real-world images

1. INTRODUCTION

The problem of human pose estimation, defined as the problem of localization of human joints, has enjoyed substantial attention in the computer vision community. one can see some of the challenges of this problem – strong articulations, small and barely visible joints, occlusions and the need to capture the context. The main stream of work in this field has been motivated mainly by the first challenge, the need to search in the large space of all possible articulated poses. Part-based models lend themselves naturally to model articulations and in the recent years a variety of models with efficient inference have been proposed The above efficiency, however, is achieved at the cost of limited expressiveness – the use of local detectors, which reason in many cases about a single part, and most importantly by modeling only a small subset of all interactions between body parts. These limitations, as exemplified, have been recognized and methods reasoning about pose in a holistic manner have been proposed but with limited success in real-world problems. In this work we ascribe to this holistic view of human pose estimation. We capitalize on recent developments of deep learning and propose a novel algorithm based on a Deep Neural Network (DNN). DNNs have shown outstanding performance on visual classification tasks and more recently on object localization. However, the question of applying DNNs for precise localization of articulated objects has largely remained unanswered. In this paper we attempt to cast a light on this question and present a simple and yet powerful formulation of holistic human poseestimation as a DNN. We formulate the pose estimation as a joint regression problem and show how to successfully cast it in DNN settings. The location of each body joint is regressed to using as an input the full image and a 7-layered generic convolutional DNN. There are two advantages of this formulation. First, the DNN is capable of capturing the full context of each body joint – each joint regressor uses the full image as a signal. Second, the approach is substantially simpler to formulate than methods based on graphical models – no need to explicitly design feature representations and detectors for parts; no need to explicitly design a

model topology and interactions between joints.

Instead, we show that a generic convolutional DNN can be learned for this problem. Further, we propose a cascade of DNN-based pose predictors. Such a cascade allows for increased precision of joint localization. Starting with an initial pose estimation, based on the full image, we learn DNN-based regressors which refines the joint predictions by using higher resolution sub-images. We show state-of-art results or better than state-of-art on four widely used benchmarks against all reported results. We show that our approach performs well on images of people which exhibit strong variation in appearance as well as articulations. Finally, we show generalization performance by cross-dataset evaluation.

2. LITERATURE SURVEY

M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection and articulated pose estimation. In CVPR, 2009. Non-rigid object detection and articulated pose estimation are two related and challenging problems in computer vision. Numerous models have been proposed over the years and often address different special cases, such as pedestrian detection or upper body pose estimation in TV footage. This paper shows that such specialization may not be necessary, and proposes a generic approach based on the pictorial structures framework. We show that the right selection of components for both appearance and spatial modeling is crucial for general applicability and overall performance of the model. The appearance of body parts is modeled using densely sampled shape context descriptors and discriminatively trained AdaBoost classifiers. Furthermore, we interpret the normalized margin of each classifier as likelihood in a generative model. Non-Gaussian relationships between parts are represented as Gaussians in the coordinate system of the joint between parts. The marginal posterior of each part is inferred using belief propagation. We demonstrate that such a model is equally suitable for both detection and pose estimation tasks, outperforming the state of the art on three recently proposed datasets.

M. Dantone, J. Gall, C. Leistner, and L. Van Gool. Human pose estimation using body parts dependent joint regressors. In CVPR, 2013. In this work, we address the problem of estimating 2d human pose from still images. Recent methods that rely on discriminatively trained deformable parts organized in a tree model have shown to be very successful in solving this task. Within such a pictorial structure framework, we ddress the problem of obtaining good part templates by proposing novel, non-linear joint regressors. In particular, we employ two-layered random forests as joint regressors. The first layer acts as a discriminative, independent body part classifier. The second layer takes the stimated class distributions of the first one into account and is thereby able to predict joint locations by modeling the interdependence and co-occurrence of the parts. This results in a pose estimation framework that takes dependencies between

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

body parts already for joint localization into account and is thus able to circumvent typical ambiguities of tree structures, such as for legs and arms. In the experiments, we demonstrate that our body parts dependent joint r'

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM:

The idea of representing articulated objects in general, and human pose in particular, as a graph of parts has been advocated from the early days of computer vision . The so called Pictorial Strictures (PSs), introduced by Fishler and Elschlager , were made tractable and practical by Felzenszwalb and Huttenlocher using the distance transform trick. As a result, a wide variety of PS-based models with practical significance were subsequently developed. The above tractability, however, comes with the limitation of having a tree-based pose models with simple binary potential not depending on image data. As a result, research has focused on enriching the representational power of the models while maintaining tractability. Earlier attempts to achieve this were based on richer part detectors . More recently, a wide variety of models expressing complex joint relationships were proposed. Yang and Ramanan [26] use a mixture model of parts. Mixture models on the full model scale, by having mixture of PSs, have been studied by Johnson and Everingham .

DisAdvantages:

Unable to predict the pose,

No accuracy in time

3.2 PROPOSED SYSTEM:

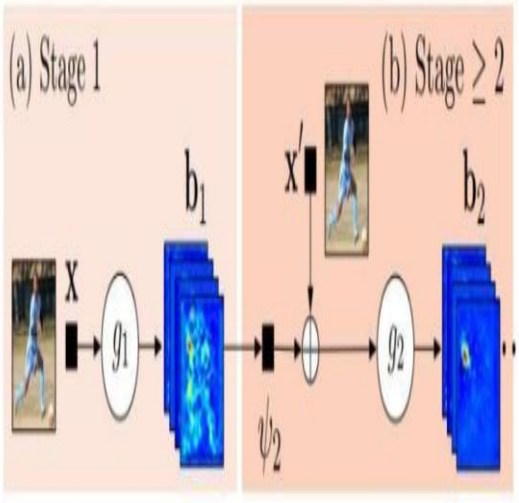
In this work, we treat the problem of pose estimation as regression, where the we train and use a function $\psi(x; \theta) \in R\ 2k$ which for an image x regresses to a normalized pose vector, where θ denotes the parameters of the model. Thus, using the normalization transformation the pose prediction $y *$ in absolute image coordinates reads $y * = N -1 (\psi(N(x); \theta))$ Despite its simple formulation, the power and complexity of the method is in ψ, which is based on a convolutional Deep Neural Network (DNN). Such a convolutional network consists of several layers – each being a linear transformation followed by a non-linear one. The first layer takes as input an image of predefined size and has a size equal to the number of pixels times three color channels. The last layer outputs the target values of the regression, in our case 2k joint coordinates.

Advantages:

Can predict the pose

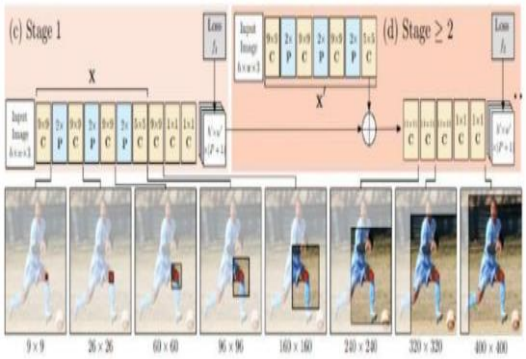
accuracy in time

4 Novel Algorithm



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

g1() and g2() predict heatmaps (belief maps in the paper). Above is a high level view. Stage 1 is the image feature computation module, and Stage 2 is the prediction module. Below is a detailed architecture. Notice how the receptive fields increase in size?



A CPM can consist of > 2 Stages, and the number of stages is a hyperparameter. (Usually = 3). Stage 1 is fixed and stages > 2 are just repetitions of Stage 2. Stage 2 take heatmaps and image evidence as input. The input heatmaps add spatial context for the next stage. (Has been discussed in detail in the paper). On a high level, the CPM refines

the heatmaps through subsequent stages. The paper used intermediate supervision after each stage to avoid the problem of vanishing gradients, which is a common problem for deep multi-stage networks.

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

Ø What data should be given as input?

Ø How the data should be arranged or coded?

Ø The dialog to guide the operating personnel in providing input.

Ø Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

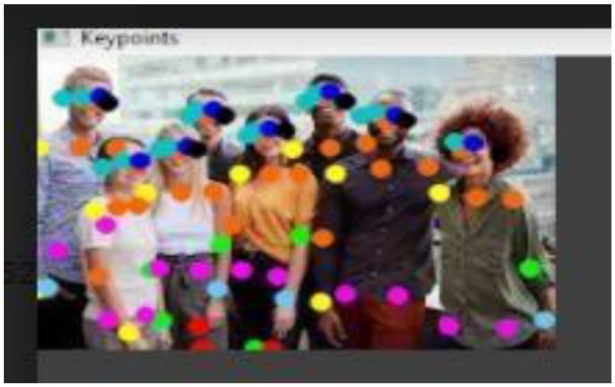
10. OUTPUT SCREENS

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intel igent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

v Convey information about past activities, current status or projections of the
v Future.
v Signal important events, opportunities, problems, or warnings.
v Trigger an action.
v Confirm an action.



CONCLUSION:

We present, to our knowledge, the first application of Deep Neural Networks (DNNs) to human pose estimation. Our formulation of the problem as DNN-based regression to joint coordinates and the presented cascade of such regressors has the advantage of capturing context and reasoning about pose in a holistic manner. As a result, we are able to achieve state-of-art or better results on several challenging academic datasets. Further, we show that using a generic convolutional neural network, which was originally designed for

classification tasks, can be applied to the different task of localization. In future, we plan to investigate novel architectures which could be potentially better tailored towards localization problems in general, and in pose estimation in particular.

REFERENCES

[1] M. Andriluka, S. Roth, and B. Schiele. Pictorial structures revisited: People detection
and articulated pose estimation. In CVPR, 2009.
[2] M. Dantone, J. Gall, C. Leistner, and L. Van Gool. Human pose estimation using body
parts dependent joint regressors. In CVPR, 2013.
[3] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and
stochastic optimization. In COLT. ACL, 2010.
[4] M. Eichner and V. Ferrari. Better appearance models for pictorial structures. 2009.
[5] M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari. Articulated human pose
estimation and search in (almost) unconstrained still images. ETH Zurich, D-ITET,
BIWI, Technical Report No, 272, 2010.
[6] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition.
International Journal of Computer Vision, 61(1):55–79, 2005.
[7] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Progressive search space reduction for
human pose estimation. In CVPR, 2008.
[8] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. Computers, IEEE Transactions on, 100(1):67–92, 1973.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate
object detection and semantic segmentation. In CVPR, 2014.
[10] G. Gkioxari, P. Arbelaez, L. Bourdev, and J. Malik. Articu- ´ lated pose estimation using
discriminative armlet classifiers. In CVPR, 2013.