Scalable Color Sketch based Search of Millions of Images

Tu Bui, University of Surrey; and John Collomosse, University of Surrey

The volume of visual data consumed on mobile devices is growing exponentially. Gestural interaction, such as sketching, provides a convenient and intuitive modality for interacting with visual content on such devices, which primarily feature touch-screens rather than keyboards. Yet despite regaining significant traction in the research community, sketch based image retrieval (SBIR) has not yet seen wide-spread adoption for visual search. Possible explanations for this include a focus primarily on shape (structure) alone in SBR, and a lack of scalability with most techniques able to index only a few thousand images within practical query times (i.e. sub-second response).

We present a live demonstration of our sketch based visual search software, capable of searching the ImageNet dataset (16 million images) in less than one second. The user is invited to draw a free-hand sketch via a tablet interface, using either the touch-screen or stylus. Matching images are displayed in real-time as the sketch develops. Uniquely our system enables search using colour sketches, and scales with state of the art accuracy over millions of images. A video of the demo software is available at https://www.youtube.com/watch?v=XSIpGCXgkLM.

Impresee: Searching in Catalogs using Photos and Sketches

Juan Manuel Barrios, Orand S.A.; and Jose Manuel Saaavedra, Orand S.A

In this demo we present a mobile application for retrieving products using hand-drawn sketches and photographs. The application is intended to be published by retail stores and design stores, allowing customers to search and buy products by drawing the shape of the desired product, or (if the user is in front of an example product) by taking a photograph with a mobile device. The sketch or photo is sent over Internet to a central server, which perform a similarity search in catalogs for different stores and returns the most relevant products.

The similarity search is based on our research on sketch-based image retrieval and efficient kNN. The sketch or photo is first classified into categories using a CNN. A visual similarity search is then performed based on sketch descriptors for sketches, or local descriptors for photos. The final result contains products from the same category and/or visually alike to the query sketch or photo.

A mobile application is publicly available to query the catalog of a Chilean retail store. We have tested an extension of this technology for store POS. We are currently working on developing a public API in order to embed our search engine into third-party mobile applications.

Robust Monocular SLAM for Augmented Reality

Haomin Liu, Zhejiang University; Jinyu Li, Zhejiang University; Guofeng Zhang, Zhejiang University; and Hujun Bao, Zhejiang University

We present a novel monocular SLAM system which can work in a large-scale scene. Our proposed non-consecutive feature tracking method not only can significantly extend the lifetime of feature tracks, but also can efficiently recognize and match the common features in different subsequences or even different videos. We also contribute an effective segment-based SfM estimation scheme, which can efficiently and globally optimize the structure and motion with limited memory space. Even for a large-scale scene, our method still can real-time detect and close the loop closure to eliminate the accumulation error and drift. Our method also has been successfully extended to a mobile device (e.g. iPhone) for realtime SLAM. For a mobile device, we further propose to incorporate IMU data to aid visual SLAM which can significantly improve the robustness in challenging cases (e.g. there are serious motion blur, large textureless regions and dynamic objects). Some AR applications will be presented to demonstrate that the proposed system outperforms the existing state-of-the-art SLAM systems.

Live demo for Semantic Image Segmentation

Shuai Zheng, University of Oxford; Sadeep Jayasumana, University of Oxford; Bernardino Romera-Paredes, University of Oxford; Philip H. S. Torr, University of Oxford

This is an online live demo for semantic image segmentation. Our work allows the computer to recognize objects in images, and also to recover the 2D outline of the objects, which is a distinctive feature of our demo compared to traditional object classification/detection demos. This work is part of a project to build augmented reality glasses for the partially sighted. Currently we have trained this model to recognize 20 classes. The demo allows you to test our algorithm on your own images.

This demo is based on our ICCV 2015 paper titled “Conditional Random Fields as Recurrent Neural Networks”. In this work, we introduce a new form of convolutional neural network that combines the strengths of Convolutional Neural Networks (CNNs) and probabilistic graphical modelling. To this end, we formulate the filter-based approximate mean-field inference as a recurrent neural network. This network, called CRF-RNN, is then plugged in as a part of a CNN to obtain a deep network that has desirable properties of both CNNs and CRFs. Importantly, our system fully integrates CRF modelling with CNNs, making it possible to train the whole deep network end-to-end with the usual back-propagation algorithm, avoiding offline post-processing methods for object delineation.

Showcasing SegNet: A Deep Encoder-Decoder Architecture for Real-Time Road Scene Segmentation

Vijay Badrinarayanan, University of Cambridge; Alex Kendall, University of Cambridge; Roberto Cipolla, University of Cambridge

SegNet is a deep convolutional network architecture designed to map input RGB images to pixel labels. It is composed of an encoder network and a decoder network which ends with a softmax classifier. The encoder network can be any convolutional neural network which is designed for classification tasks such as the VGG16 network. The key element of SegNet is the decoder network which maps the lossy encoder representation to pixel labels. A decoder in the decoder network upsamples its input feature map(s) by using the stored indices of the max feature locations in the corresponding encoder feature map. It then convolves these feature maps with a trainable decoder filter bank. This decoding technique has the advantage that it reduces the number of parameters in the decoder network by a significant margin as compared to other recent architectures proposed for segmentation which learn to upsample and consequently enables endtoend training of the whole architecture. We add here that such a decoding technique was proposed for layerwise unsupervised learning of features in a deep network by Ranzato et al. The difference to SegNet is that the decoders were only instrumental to learning the encoders and were discarded soon after, and secondly, endtoend learning of the network was not their goal. The elements of the SegNet architecture are described in Badrinarayanan, Vijay, Ankur Handa, and Robert Cipolla. "SegNet: A Deep Convolutional EncoderDecoder Architecture for Robust Semantic PixelWise Labelling." arXiv preprint arXiv:1505.07293 (2015). The demo however will use a more elaborate form of SegNet which is trained endtoend using SGD.

Automatic Segmentation of Rectangular Patches in 3D Point Clouds

William Nguatem, Bundeswehr University Munich

We propose a fully automatic and fast technique to generate consistent rectangular patches from 3D point clouds of out-door scenes. Our algorithm includes RANSAC for robust estimation and stratifies the sample space of rectangular patches using a spatial partitioning routine. Using Bayesian Nonparametric technique, and without making any assumption of the nature (sparse, dense), origin (LIDAR, image-matching) of the input point clouds, we build consistent clusters of planar patches. We present also the feasibility of our algorithms for the online and interactive segmentation of point clouds from an RGB-D Depth Sensor (e.g. Kinect).

FlowNet: Real-time Optical Flow Estimation with Convolutional Networks

Alexey Dosovitskiy, University of Freiburg; Philipp Fischer, University of Freiburg; Eddy Ilg, University of Freiburg; Philip Haeusser, Technical University of Munich; Daniel Cremers, Technical University of Munich; and Thomas Brox, University of Freiburg

For the first time we demonstrate accurate optical flow estimation with a convolutional network, running in real time on a laptop. Convolutional neural networks (CNNs) have recently been very successful in a variety of computer vision tasks, especially on those linked to recognition. Optical flow estimation has not been among the tasks where CNNs were successful. We constructed CNNs which are capable of solving the optical flow estimation problem as a supervised learning task. We demonstrate and compare two architectures: a generic architecture and another one including a layer that correlates feature vectors at different image locations. The networks were trained on a synthetic Flying Chairs dataset. The demo shows that networks trained on this unrealistic data still generalize very well to realistic data, achieving competitive accuracy at frame rates of roughly 10 fps.

ALIEN 2.0: The Infinite Memory

Federico Pernici, University of Florence; and Alberto Del Bimbo, University of Florence

Visual data is massive, is growing faster than our ability to store or index it and the cost of manual annotation is critically expensive. Effective methods for unsupervised learning are of paramount need. A possible scenario is that of considering visual data coming in the form of streams. In dynamically changing and non-stationary environments, the data distribution can change over time yielding the general phenomenon of concept drift which violates the basic assumption of traditional machine learning algorithms (iid).

This demo presents our recent results in learning an instance level object detector from a potentially infinitely long video-stream (i.e. YouTube). This is an extremely challenging problem largely unexplored, since a great deal of work has been done on learning under the iid assumption. Our approach starts from the recent success of long term object tracking extending our previously developed and demonstrated method (ALIEN).

The novel contribution is the introduction of an online appearance learning procedure based on an incremental condensing strategy which is shown to be asymptotically stable. Asymptotic stability evidence will be interactively evaluated by attendants based on a real time face tracking application using webcam or YouTube data.

Anisotropic Reflectance Rendering of Noh-Kimono Costumes in Dynamic Lighting Environments with Bonfire

Shiro Tanaka, Ritsumeikan University; Wataru Wakita, Hiroshima City University; and Hiromi T. Tanaka, Ritsumeikan University

The Noh (traditional masked dance-drama) is one of the traditional arts in Japan. Among this, there is one called Takigi (firewood) Noh which is performed by burning firewood around the stage after sunset. Noh costume of gold brocades shines and flickers beautifully and intricately. In this work, we propose a real-time bidirectional texture function (BTF) and image-based lighting (IBL) rendering of the Takigi Noh based on reflectance analysis. Firstly, we observe fabrics of the Noh costume by omnidirectional anisotropic reflectance measurement system called Optical Gyro Measuring Machine (OGM), and we modeled the BTF data of the weave pattern based on multi-illuminated High Dynamic Range (HDR) image analysis. Secondly, from the BTF image data of the fabric, anisotropic parameters of Ashikhmin reflectance model are estimated and the incident direction of each weave pattern will be determined using Importance Sampling. Based on the determined incident direction, individual pixels can be obtained from the dynamic light texture. Finally, we showed the rendering of the anisotropic reflectance of the Noh costume and dancing Noh player wearing the costume under dynamic lighting by Bonfire of the Takigi based on IBL.

The Menpo Project

Patrick Snape, Imperial College London; and Epameinondas Antonakos, Imperial College London

The Menpo Project (http://www.menpo.org/) is a BSD-licensed set of tools and software designed to provide an end-to-end pipeline for collection and semi-automatic annotation of image and 3D mesh data. In particular, the Menpo Project provides tools for annotating images and meshes of an object class with a sparse set of fiducial markers that we refer to as landmarks. These landmarks are useful in a variety of areas in Computer Vision including object detection, deformable modelling and tracking.

We are actively developing and contributing to the state-of-the-art in deformable modelling. Of most interest to the Computer Vision is the fact that The Menpo Project contains completely open source implementations of a number of state-of-the-art algorithms for object detection and deformable model building. Moreover, Menpo's online Landmarker (https://www.landmarker.io/) is an ideal tool for annotating images of any object class. Even though The Menpo Project is designed to deal with any object class, our demo will consist of face-specific tasks, such as real-time face detection and tracking, semi-automatic annotation of facial images and short tutorials on building and fitting deformable models.

Expresso: A User-Friendly GUI for Designing, Training and Exploring Convolutional Neural Networks

Santosh Ravi Kiran Sarvadevabhatla, Indian Institute of Science; and Venkatesh Babu R, Indian Institute of Science

With a view to provide a user-friendly interface for designing, training and developing deep learning frameworks, we have developed Expresso (http://val.serc.iisc.ernet.in/expresso/), an open-source GUI tool written in Python. Expresso is built atop Caffe, the opensource, prize-winning framework popularly used to develop Convolutional Neural Networks. Expresso provides a convenient wizard-like graphical interface with a contextual help interface to guide the user through various common scenarios -- data import, construction and training of deep networks, performing various experiments, analyzing and visualizing the results of these experiments. A set of detailed and illustrated tutorials in text and narrated video formats are also provided. We believe Expresso's flexibility and ease of use will come in handy to researchers, newcomers and seasoned alike, in their explorations related to deep learning.

Interactive Action Recognition based on Motion Capture Data

Fabrizio Natola, University of Rome; Valsamis Ntouskos, University of Rome; Fora Pirri, University of Rome; and Marta Sanzari, University of Rome

In this demo we will demonstrate the ability of our algorithm, presented at the main conference, to recognize different types of actions performed by a subject based on his/her movements. The recognition is performed on the data provided by a real-time motion capture (MoCap) system used by the subject.

In particular, we will equip attendants willing to participate to the demo with the sensors of the Xsens MVN Biomech Awinda system. The MVN Biomech system is an IMU-based Motion Capture (MoCap) system able to capture the movements of a subject in real-time. The participants will be encouraged to perform randomly different actions and our system will recognize the action being performed.

The motion sequence captured by the MoCap system is provided to our algorithm in realtime. First, the MoCap feature representation is computed. This representation is then used to compare the performed action with the set of actions stored in our training dataset, composed of MoCap sequences taken from the HDM05 dataset, allowing the classification of the input sequence.