

## Let the Image Speak (Real time image captioning based object detection and localization)

<b>PROJECT ID</b>   CS15-8		<b>TYPE</b>   [ X ] New [ ] Continuing		<b>START DATE</b>   July 2015											
<b>PROJECT LEAD/PARTICIPANTS</b>   Prof. Serkan Kiranyaz, Dr. Iftikhar Ahmad, Muhammad Waris															
<b>DESCRIPTION</b>   Automatic image content description is a vital problem in artificial intelligence that connects computer vision and natural language processing. The primary challenge towards this goal is in the design of a multi-model approach that is rich enough to aim simultaneously about contents of images and their representation in terms of words or sentences. We present a multi-model approach based on a deep learning architecture that combines recent advances in computer vision such as; salient object proposal prediction, object detection to generate natural sentences describing an image. Leveraging recent advances in recognition of objects, their attributes and locations, allows us to drive natural language generation systems, however they are limited in their expressivity. Moreover, current object detection methods still suffer various problems in localization and processing time that make them unreliable and inadequate as they are still slow at test time. We shall attack this problem within three main research scenarios, i.e. category independent saliency based object proposals with two distinct classification schemes, Learning with a large but a single Operational Neural Network (ONN) and 'Divide and Conquer' Learning with ensembles of simple but 'expert' ONNs, each of which dedicated for a subset of the dataset with a certain homogeneity.															
<b>EXPERIMENTAL PLAN</b>   During the project period the team will work to: <ul style="list-style-type: none"> <li>• Object Proposal (Automatic part based salient object extraction)</li> <li>• Perform deep learning by evolutionary Operational Neural Network (ONN) to "learn" and "mimic" the human visual system.</li> <li>• Attack the learning problem within two research scenarios: Learning with a large but a single ONN and 'Divide and Conquer' learning with ensembles of simple but 'expert' ONNs for object detection.</li> <li>• Exploring other computer vision task such as face recognition, action recognition and pose estimation that can benefit in better prediction.</li> </ul>															
<b>RELATED WORK</b>   We are one of the leading research groups in providing salient object extraction. Our recent publication " <i>Automatic Object Segmentation by Quantum Cuts</i> " won the "IBM Best Paper Award" in IEEE ICPR'14, which is the largest and perhaps the most prestigious conference in computer vision, machine intelligence and pattern recognition. Recently our team participated in "The 2013 Face Recognition Evaluation in Mobile Environment" showed competitive results compared to other participant. We are very active in the field of Evolutionary ANNs where the work published in Neural Networks journal has become one of the most-read and the most-cited papers in the Journal's history.															
<b>HOW OURS IS DIFFERENT</b>   Several related works in this problem rely on holistic scene understanding in which the scene type, objects and their spatial support in the image is inferred. However, these methods rely on a large amount of training data to capture the variety in possible object appearances, and are often expensive at test time due to their non-parametric nature. However, we target the high-level goal of annotating the contents of images based salient regions or segments of images and study the multimodal correspondence between words and images. The idea is to correctly label scenes, objects and regions with a fixed set of categories, while our focus is on richer and higher-level region descriptions. The proposed approaches can also be used in text to image search in large scale image retrieval systems.			<b>MILESTONES FOR YEAR</b>   <p><i>3 months:</i> Saliency based object proposal for fast and object detection and better localization framework, Realization of stand-alone operational neural networks(ONNs)</p> <p><i>6 months:</i> Realization of ensemble network of ONNs.</p> <p><i>9 months:</i> Extension of both approaches to achieve state-of-art results. For this purpose the implementation efficient large-scale ensembles of intelligent systems over the Grid/Cloud environment will be performed.</p> <p><i>12 months:</i> The testing and evaluation of both approaches in order to maximize the accuracy for several use-case scenarios. Realization of the final system.</p>												
<b>DELIVERABLES</b>   <ol style="list-style-type: none"> <li>1. Category independent efficient saliency based object proposals</li> <li>2. Realization of stand-alone operational neural networks (ONNs)</li> <li>3. Realization of the Self-Organized Network of ONNs</li> <li>4. Cloud computing and efficient utilization of system resources</li> <li>5. Realization and integration into the final system.</li> </ol>			<b>BUDGET FOR YEAR</b>   <table> <tr> <td>Student Salary</td> <td>€55,000</td> </tr> <tr> <td>Salary Overhead:</td> <td>€27,500</td> </tr> <tr> <td>Travel</td> <td>€7,500</td> </tr> <tr> <td>Overhead</td> <td>€10,000</td> </tr> <tr> <td><b>Total</b></td> <td><b>€100,000</b></td> </tr> </table>			Student Salary	€55,000	Salary Overhead:	€27,500	Travel	€7,500	Overhead	€10,000	<b>Total</b>	<b>€100,000</b>
Student Salary	€55,000														
Salary Overhead:	€27,500														
Travel	€7,500														
Overhead	€10,000														
<b>Total</b>	<b>€100,000</b>														
<b>ECONOMICS</b>   Humans can point out and describe an immense amount of details about the visual scene with just quick glance at an image. Automatic image content description can be used in wide area of application. Idea of revolutionizing machines ability for content description is an active research field from last 2-3 years. The proposed method is expected to lead to considerable gains in terms of speed and accuracy.															
<b>POTENTIAL MEMBER COMPANY BENEFITS</b>   CVDI members use advance machine learning algorithms to make the most efficient learning, decision making, classification and recognition. Such accomplishments will also complement ongoing CVDI projects that employ any form of learning and intelligent system implementation. Large scale image databases are becoming common in various industries and companies. Such novel approaches may lead to crucial advantages in visual understanding, analysis, retaining and acquiring customers and of course increasing revenues.															
<b>PROGRESS TO DATE</b>   A rich set of algorithms were developed in the past 10 years. So far, we have tested the initial version of ONNs and observed very satisfactory performance level. We also conducted small-scale analysis and evaluation of face spoofing.															
<b>KNOWLEDGE TRANSFER TARGET DATE</b>   12 months															