Challenges in processing and knowledge discovery in specifications of scientific resources

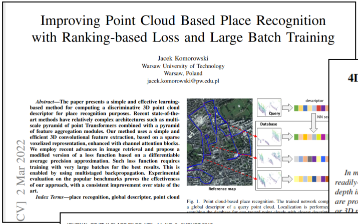
Witold Taisner

Poznan Supercomputing and Networking Center

TNC2022

What's wrong with doing science?

What's wrong with doing science?



Improving Point Cloud Based Place Recognition with Ranking-based Loss and Large Batch Training

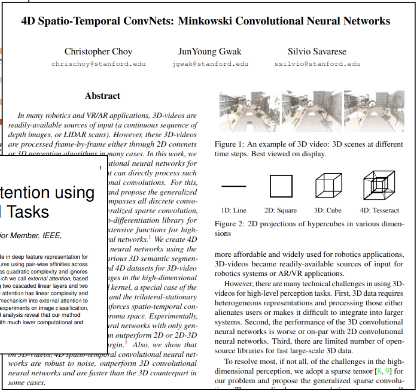
Jacek Komorowski
Warsaw University of Technology
Warsaw, Poland
jacek.komorowski@pw.edu.pl

CV] 2 Mar 2022

Abstract—The paper presents a simple and effective learning-based method for computing a discriminative 3D point cloud descriptor for place recognition purposes. Recent state-of-the-art methods have relatively complex architectures such as multi-scale pyramid of point Transformers combined with a pyramid of feature aggregation modules. Our method uses a simple and efficient 3D convolutional feature extraction, based on a sparse voxelized representation, enhanced with channel attention blocks. We employ recent advances in image retrieval and propose a modified version of a loss function based on a differentiable average precision approximation. Such loss function requires training with very large batches for the best results. This is enabled by using multistaged backpropagation. Experimental evaluation on the popular benchmarks proves the effectiveness of our approach, with a consistent improvement over state of the art.

Index Terms—place recognition, global descriptor, point cloud

Fig. 1. Point cloud-based place recognition. The trained network comp
a global descriptor of a query point cloud. Localization is performed



4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks

Christopher Choy chrischoy@stanford.edu
JunYoung Gwak jgwak@stanford.edu
Silvio Savarese ssilvio@stanford.edu

Abstract

In many robotics and VR/AR applications, 3D-videos are readily-available sources of input (a continuous sequence of depth images, or LIDAR scans). However, these 3D-videos are processed frame-by-frame either through 2D convnets or 3D perception algorithms in many cases. In this work, we … utional neural networks for … t can directly process such … onal convolutions. For this, … and propose the generalized … mpasses all discrete convo- … eralized sparse convolution, … o-differentiation library for … xtensive functions for high- … al networks.[1] We create 4D … neural networks using the … rious 3D semantic segmen- … d 4D datasets for 3D-video … ges in the high-dimensional … kernel, a special case of the … and the trilateral-stationary … forces spatio-temporal con- … roma space. Experimentally, … ral networks with only gen- … n outperform 2D or 2D-3D … rgin.[2] Also, we show that on 3D-videos, 4D spatio-temporal convolutional neural networks are robust to noise, outperform 3D convolutional neural networks and are faster than the 3D counterpart in some cases.

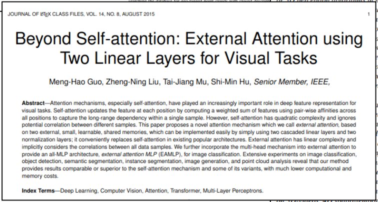
Figure 1: An example of 3D video: 3D scenes at different time steps. Best viewed on display.

Figure 2: 2D projections of hypercubes in various dimensions

more affordable and widely used for robotics applications, 3D-videos became readily-available sources of input for robotics systems or AR/VR applications.

However, there are many technical challenges in using 3D-videos for high-level perception tasks. First, 3D data requires heterogeneous representations and processing those either alienates users or makes it difficult to integrate into larger systems. Second, the performance of the 3D convolutional neural networks is worse or on-par with 2D convolutional neural networks. Third, there are limited number of open-source libraries for fast large-scale 3D data.

To resolve most, if not all, of the challenges in the high-dimensional perception, we adopt a sparse tensor [8, 9] for our problem and propose the generalized sparse convolu-



JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015 1

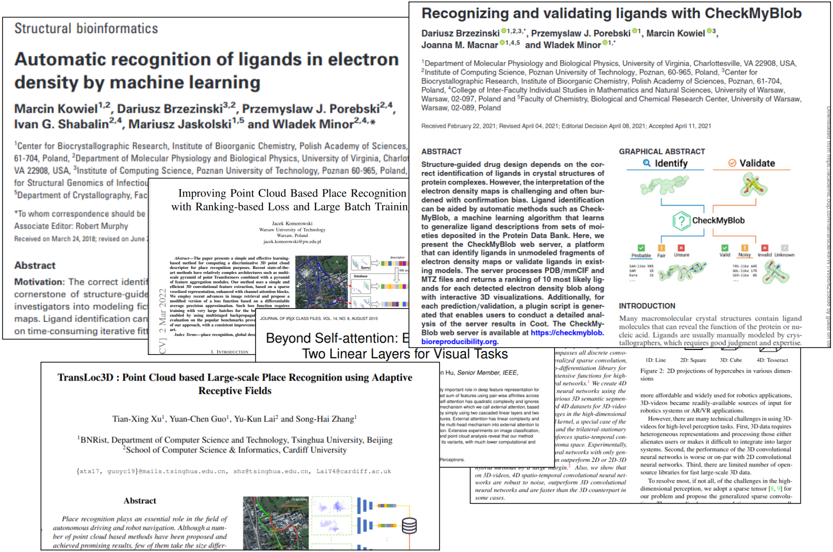
Beyond Self-attention: External Attention using Two Linear Layers for Visual Tasks

Meng-Hao Guo, Zheng-Ning Liu, Tai-Jiang Mu, Shi-Min Hu, *Senior Member, IEEE,*

Abstract—Attention mechanisms, especially self-attention, have played an increasingly important role in deep feature representation for visual tasks. Self-attention updates the feature at each position by computing a weighted sum of features using pair-wise affinities across all positions to capture the long-range dependency within a single sample. However, self-attention has quadratic complexity and ignores potential correlation between different samples. This paper proposes a novel attention mechanism which we call *external attention*, based on two external, small, learnable, shared memories, which can be implemented easily by simply using two cascaded linear layers and two normalization layers; it conveniently replaces self-attention in existing popular architectures. External attention has linear complexity and implicitly considers the correlations between all data samples. We further incorporate the multi-head mechanism into external attention to provide an all-MLP architecture, *external attention MLP* (EAMLP), for image classification. Extensive experiments on image classification, object detection, semantic segmentation, instance segmentation, image generation, and point cloud analysis reveal that our method provides results comparable or superior to the self-attention mechanism and some of its variants, with much lower computational and memory costs.

Index Terms—Deep Learning, Computer Vision, Attention, Transformer, Multi-Layer Perceptrons.

What's wrong with doing science?



Why is EOSC so cool?

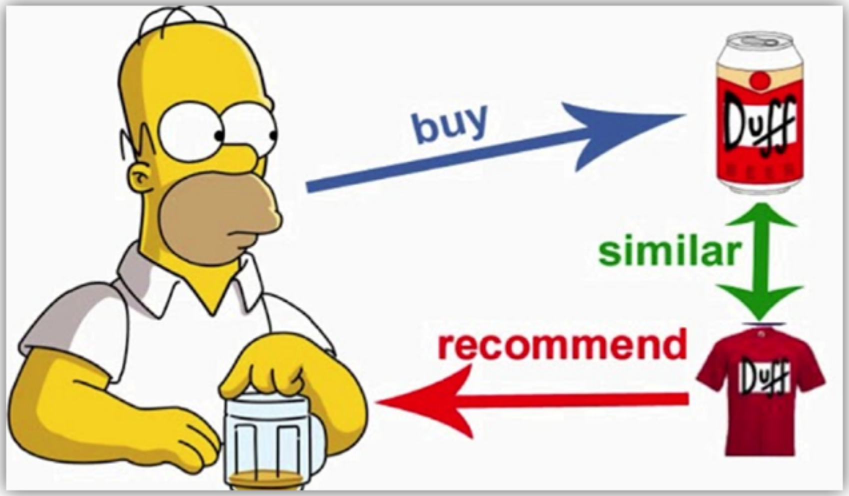
European Open Science Cloud

One system to rule them all, one system to find them, One system to bring them all, and in the references bind them; In the Land of Citations where the shadows lie.

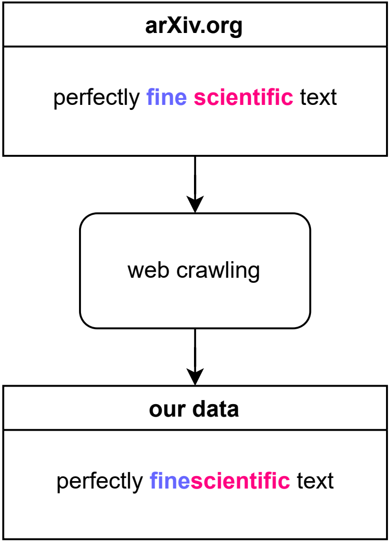
EUROPEAN OPEN SCIENCE CLOUD

https://eosc-portal.eu/

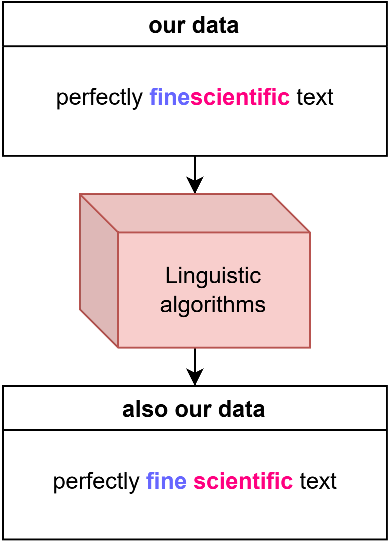
EOSC Recommender System



Linguistic knowledge

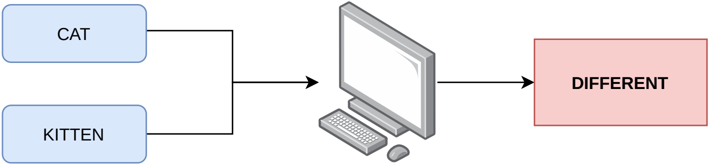


Linguistic knowledge

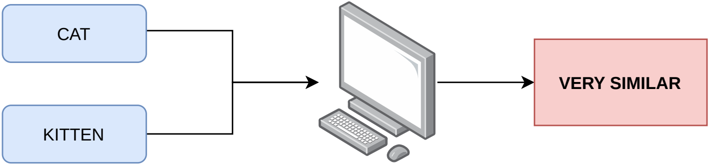


Deep learning

How it looks:



How it should look:



Contribute to open science!