LET'S HAVE A LOOK INTO YOLOR

AI TEAM 3
TUNSA

APRIL 10, 2022

YOLOR STANDS FOR : YOU ONLY LEARN ONE REPRESENTATION

" UNIFIED NETWORKS FOR MULTIPLE TASKS"

AI TEAM 3
TUNSA

APRIL 10, 2022

AUTHOR
PAPER

You Only Learn One Representation: Unified Network for Multiple Tasks

Chien-Yao Wang[1], I-Hau Yeh[2], and Hong-Yuan Mark Liao[1]
[1]Institute of Information Science, Academia Sinica, Taiwan
[2]Elan Microelectronics Corporation, Taiwan
kinyiu@iis.sinica.edu.tw, ihyeh@emc.com.tw, and liao@iis.sinica.edu.tw

Abstract

People "understand" the world via vision, hearing, tactile, and also the past experience. Human experience can be learned through normal learning (we call it explicit knowledge), or subconsciously (we call it implicit knowledge). These experiences learned through normal learning or subconsciously will be encoded and stored in the brain. Using these abundant experience as a huge database, human beings can effectively process data, even they were unseen beforehand. In this paper, we propose a unified network to encode implicit knowledge and explicit knowledge together, just like the human brain can learn knowledge from normal learning as well as subconsciousness learning. The unified network can generate a unified representation to simultaneously serve various tasks. We can perform kernel space alignment, prediction refinement, and multi-task learning in a convolutional neural network. The results demonstrate that when implicit knowledge is introduced into the neural network, it benefits the performance of all tasks. We further analyze the implicit representation learnt from the proposed unified network, and it shows great capability on catching the physical meaning of different tasks. The source code of this work is at : https://github.com/WongKinYiu/yolor.

1. Introduction

Figure 1: Human beings can answer different questions from the same input. Our aim is to train a single deep neural network that can serve many tasks.

inition of how implicit learning operates and how to obtain implicit knowledge. In the general definition of neural networks, the features obtained from the shallow layers are often called explicit knowledge, and the features obtained from the deep layers are called implicit knowledge. In this paper, we call the knowledge that directly correspond to observation as explicit knowledge. As for the knowledge that is implicit in the model and has nothing to do with observation, we call it as implicit knowledge.

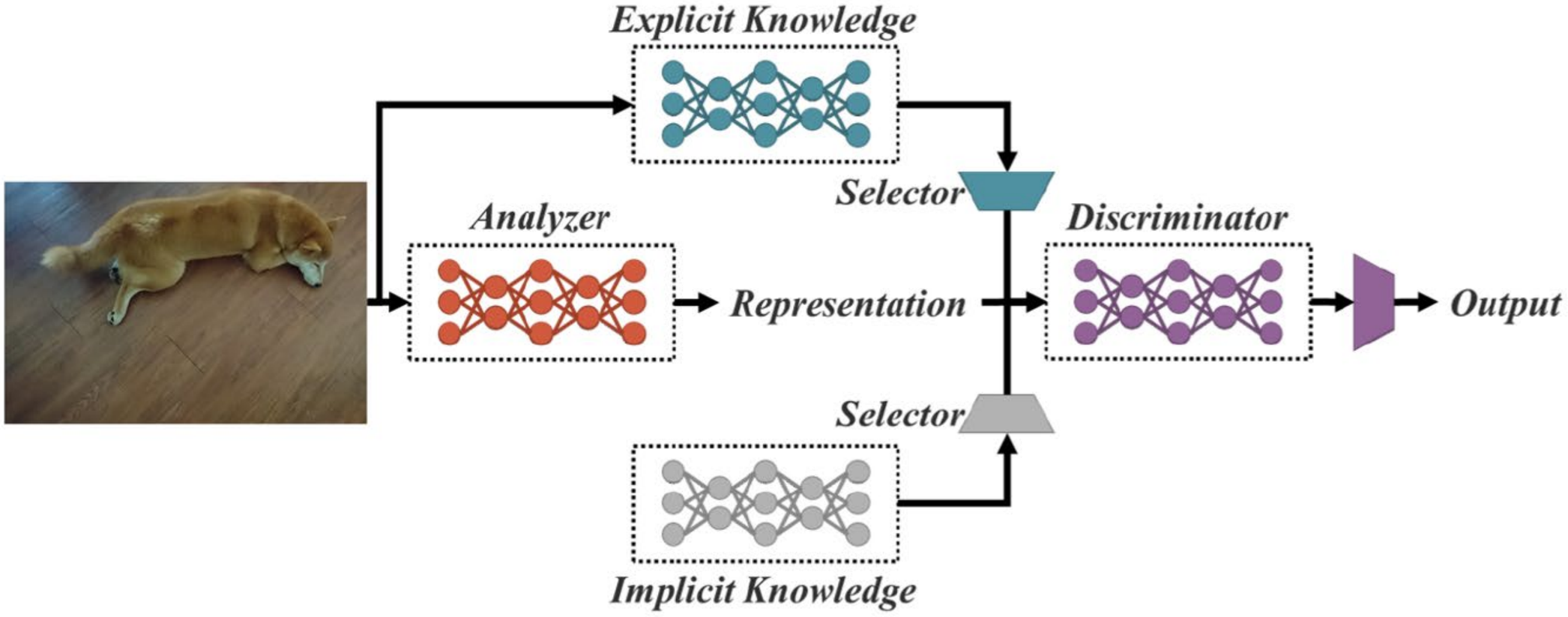
We propose a unified network to integrate implicit knowledge and explicit knowledge, and enable the learned model to contain a general representation, and this general representation enable sub-representations suitable for various tasks. Figure 2.(c) illustrates the proposed unified network architecture.

The way to construct the above unified networks is to combine compressive sensing and deep learning, and the main theoretical basis can be found in our previous work

So say that you and I had to look at a person for example, as humans, we can recognize that it's a person quite easily. Well so can a Convolutional Neural Network (CNN).
However, both you and I can also recognize where the hands and legs are, what the person is wearing, where they are located, is it in a room or outside, are they standing or jumping or playing Fortnight.
A CNN can only do one thing and do that one task robustly, but fail miserably at other tasks. Why is this? Well it comes down to 2 things:

- **Explicit, &**
- **Implicit Knowledge.**

THE UNIFIED MODEL



IMPLEMENTATION

EXPLICIT DEEP LEARNING

The authors essentially used Scaled YOLOv5 CSP

IMPLICIT DEEP LEARNING

- Manifold Space Reduction,
- Kernel Alignment, and
- More Functions.

MATHEMATICAL MODELS

CONVENTIONAL NETWORK MODEL

If we had to model a Conventional N etwork it would look something like this :

$$y = f_\theta(\mathbf{x}) + \epsilon$$

x : the observation, for example, you see a dog.
Theta : the set of parameters of a neural network,
f_θ : the operation of the neural network,
ϵ : the error term.
=> The goal is to minimize the error to make f(x) with respect to θ as close to the target y as possible.

UNIFIED NETWORKS

We can now expand our equation to incorporate the implicit model with g theta and the explicit error from observation x together with the implicit error from z which they term the hidden code : the representation of compressed data that makes up the implicit knowledge.

$$y = f_\theta(\mathbf{x}) + \epsilon + g_\phi(\epsilon_{ex}(\mathbf{x}), \epsilon_{im}(\mathbf{z}))$$

WE CAN FURTHER SIMPLIFY THE EQUATION TO THIS :.

$$y = f_\theta(\mathbf{x}) \star g_\phi(\mathbf{z})$$

With the $\star$ representing some possible operators like addition or concatenation that combines f and g or, rather the explicit with the implicit models.

COMPARISON OF THE STATE OF THE ART ALGORITHMS

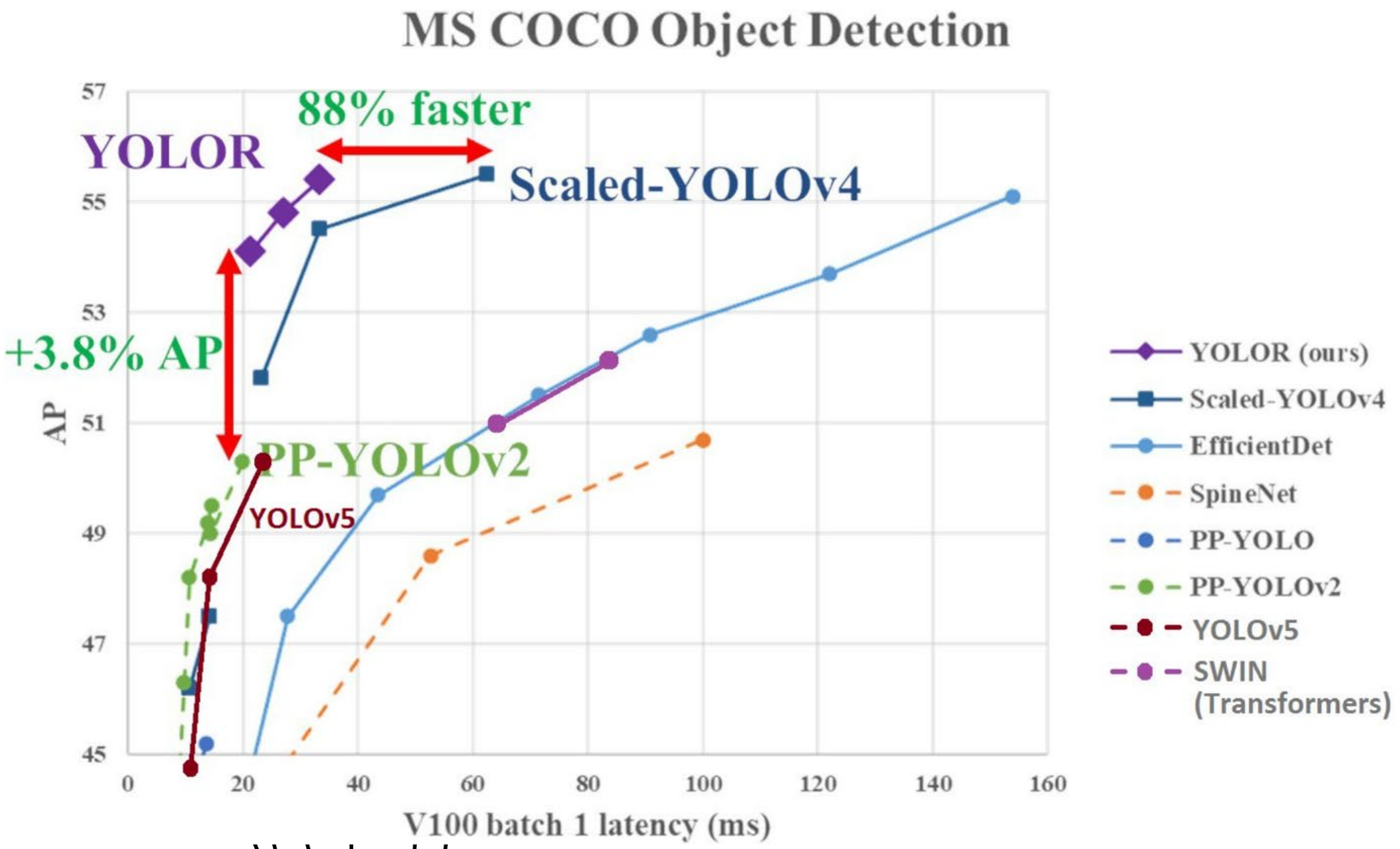
(Object Detection, image classifications, and Feature Embedding* trained on the MS-COCO Dataset)

Method	pre.	seg.	add.	AP^{test}	AP^{test}_{50}	AP^{test}_{75}	FPS^{V100}
YOLOR (ours)				55.4%	73.3%	60.6%	30
ScaledYOLOv4 [15]				55.5%	73.4%	60.8%	16
EfficientDet [13]	✓			55.1%	74.3%	59.9%	6.5
SwinTransformer [10]	✓	✓		57.7%	–	–	–
CenterNet2 [26]	✓		✓	56.4%	74.0%	61.6%	–
CopyPaste [6]	✓	✓	✓	57.3%	–	–	–

=> You can see that in terms of **accuracy** YoloR is comparable but where it shines, is in its **frame rate** : how fast an object detection model processes and generates the desired output.

*: feature embedding appears to be a synonym for feature extraction, feature learning etc. I.e. a form of embedding/dimension reduction (the goal may not be a lower dimensional representation but one of equal dimensionality, but more meaningfully expressed)

IF YOU LOOK AT THE TEST RUN BY THE LEGEND ALEXEY BOCHKOVSKIY, THEY SHOW AN 88% IN IMPROVEMENT IN SPEED WHEN COMPARED TO SCALED YOLOV4 AND A 3.8% IMPROVEMENT IN AVERAGE PRECISION COMPARED TO PP-YOLOV2.



CONCLUSION

In summary you can probably understand why the title of this paper is called You Only Learn One Representation (YOLOR) and then the second half of the title which is Unified Network for Multiple Tasks (object detection as an example).

You have also learnt how the integration of implicit knowledge along with explicit knowledge can prove very effective for multi-task learning(a machine learning approach in which we try to learn multiple tasks simultaneously, optimizing multiple loss functions at once) under a single model architecture.

THANK YOU

AI TEAM 3
TUNSA

APRIL, 10 2022