Research Report: Hard Tasks and Mixed Data Inclusion in Weak Supervision Search for Beyond Standard Model Particles

Duc Nguyen*

(Dated: August 28, 2024)

In this report, I will present my results in replicating the paper [1], which applied weak supervision methods (CWoLA, transfer learning, meta-transfer learning) to identify signal from background jet images. In general, similar to the paper, I found that transfer learning results in a significant enhancement in signal detection abilities. However, meta-transfer learning with and without hard tasks didn't result in an improvement. Finally, I will discuss the prospects of incorporating more jet information, such as Les Houches angularity and jet mass, into the neural network model.

I. INTRODUCTION

Colliders, such as the Large Hadron Collider (LHC) at CERN, are an important tool to understand particle physics. In colliders, particles are accelerated and collide with each other at high energies, producing byproducts during the process. At these energies, we can learn useful information about the composition and interactions of fundamental particles. One notable discovery that colliders facilitate is the Higgs boson particle in 2012 [2], which contributes to our framework of the Higgs field that explains the generation of mass in some particles. A current theoretical problem that colliders are exploring is whether new particles beyond the Standard Model exist, since the limitations of the Standard Model and the astronomical evidence for dark matter suggest that new particles might be present in nature.

To search for new particles, we need to identify the differences between a snapshot of a collision that hints at a new particle (signal event) and a background event, which is known as the classification problem. An effective strategy is to employ a deep learning model that can learn how to classify the events through simulations and apply the knowledge to a real search. Since such particles haven't been discovered yet, we need the model to be able to learn the difference between the signal and background across multiple new physics scenarios, and the model must work on mixed background and signal events to resemble actual data collected from the LHC.

My work used weak supervision methods (CWoLa, transfer learning, meta-transfer learning), which work with mixed samples to tackle the classification problem. I focus on jet images, which are layouts of energy depositions from collimated beams of hadrons resulting from pp collisions. In the **Exploration** section, I will also use two other measurable properties of jets: Les Houches angularity and mass as an attempt to improve the performance of the neural network.

* ducnguyenmanh@uchicago.edu; University of Chicago.

II. DATA EXTRACTION

Preprocessing the data followed the paper [1], which includes two major steps:

1. Generating events: For background events, $pp \to jj$ channel was considered. MadGraph 2.7.0 pipeline, which also includes Pythia hadronization and Delphes detector simulation, was used. For signal events, two scenarios were considered: direct decay (DD) and indirect decay (ID). For each scenario, 7 benchmarks were used, corresponding to a different value of the parameter Λ_D that governs the dark interaction processes.

2. Image processing and imposing cuts: The η and ϕ coordinate values, along with transverse momentum P_t were used to construct jet images. The images were centered based on the jets' coordinate values recorded in the event output file, rotated, and flipped. This process ensures that the neural network does not pick up spatial features. I separated the signal region (SR) events from the sideband region (SB) events for training using the dijet invariant mass M_{jj}. Signal region was defined as the region where $4700 \leq M_{jj} \leq 5500$ GeV, and the sideband region was where $4400 \leq M_{jj} \leq 4700$ GeV and $5500 \leq M_{jj} \leq 5800$ GeV. With this cut, the SR and SB region contained approximately equal number of background events, but the signal region contains more signal events. This fact is crucial to the success of the deep learning training. For each event, three different resolution (25x25, 50x50, and 75x75) images were generated.

III. DEEP LEARNING

A. Architecture

The neural network's architecture used throughout CWoLa, transfer learning and meta-transfer learning, is depicted in Fig. **1**. Exact details of the layers can be found in paper [1]. Training was performed on Keras with Tensorflow backend.

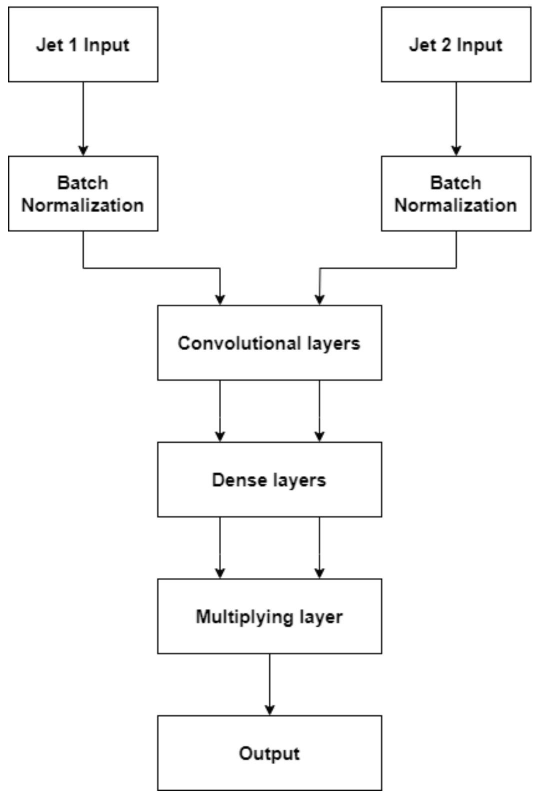

FIG. 1. The neural network architecture used for CWoLa, transfer learning and meta-transfer learning.

B. CWoLa

In CWoLa, the neural network is trained on events from the signal and sideband region rather than on pure signal and background. The mixing of signal and background events resembles the actual data that the neural network would observe in colliders if signal events are present. It can be proved that the resulting neural network is, in theory, optimized to distinguish between signal and sideband events [3].

To train a classifier, 25k background events in the SR region were used in each epoch. The amount of background events in the SB region was scaled based on the ratio of events between the regions in the initial background histogram. The same procedure was used to determine the number of signal events in the SB region with respect to those in the SR region. The number of signal events in the SR region varied, representing different significance values, which was calculated by the formula used in [1]:

$$\sigma = \sqrt{2\left((N_s + N_b)\log\left(\frac{N_s}{N_b} + 1\right) - N_s\right)} \tag{1}$$

where N_s and N_b are the number of signal and background events in the signal region, respectively.

After training, the neural network was evaluated on 20k pure signal and background samples from the SR region. To calculate the significance after training, the receiver operating characteristic curve (ROC) was used to extract signal efficiency ε_s value corresponding to different background efficiency ε_b values. The significance was then recalculated as:

$$\sigma = \sqrt{2\left((\varepsilon_s N_s + \varepsilon_b N_b)\log\left(\frac{\varepsilon_s N_s}{\varepsilon_b N_b} + 1\right) - \varepsilon_s N_s\right)} \tag{2}$$

C. Transfer learning

Transfer learning involves training on relevant tasks and transferring the knowledge to a target task. By making the neural network learn how to extract important features from the jet images from similar signal/background recognition tasks, the resulting network only needs to be fine-tuned from a smaller set of target data. In practice, my neural network was first trained on 115k background and 115k pure signal events from other benchmarks except for the target one. In the fine-tuning process, the neural network learned from mixed signal and background events similar to CWoLa. The only difference is that the convolutional layers, which had been trained before during pretraining, kept their weights throughout finetuning, and only the dense layers' weights were reinitialized and trained.

The results are presented in Fig. **2**. A neural network provides a better result when the significance after neural network cut is greater than before, which means that the curve rises above the black dotted line. In the CWoLa case, the neural network improvement happens when the significance is above around 3σ. Transfer learning is able to both provide better significance after neural network cut and a smaller threshold where the neural network produces an improvement in every resolution and every background efficiency.

D. Meta-transfer learning

Meta-transfer learning is a possible method to improve the results of transfer learning. Meta-transfer learning works by letting the neural network learn, in addition to how to extract important features, the relative importance of each feature in the jet images. This is achieved by adding scaling and shifting weights to the convolutional layers, and training these weights on additional batches of data from other benchmarks in the meta-learning phase. After training, the neural network can be evaluated to identify when to stop training and which benchmarks have the lowest accuracies.

An optional phase called hard tasks training can be performed using those benchmarks. This additional training helps the neural network "learn through hardships", meaning that the neural network will be able to recognize slight feature differences between the hard benchmarks and the background. Meta-transfer learning

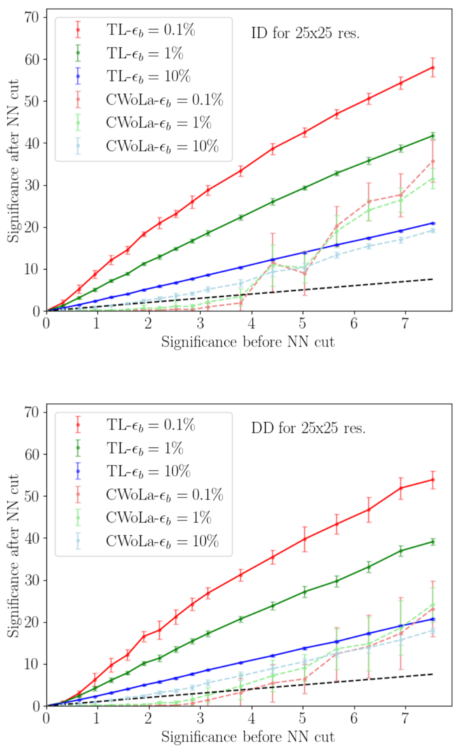

FIG. 2. CWoLa and transfer learning results for the ID10 and DD10 scenario, 25x25 resolution. Here ID refers to indirect decay, DD refers to direct decay, and 10 refers to the case where $\Lambda_D = 10$ GeV.

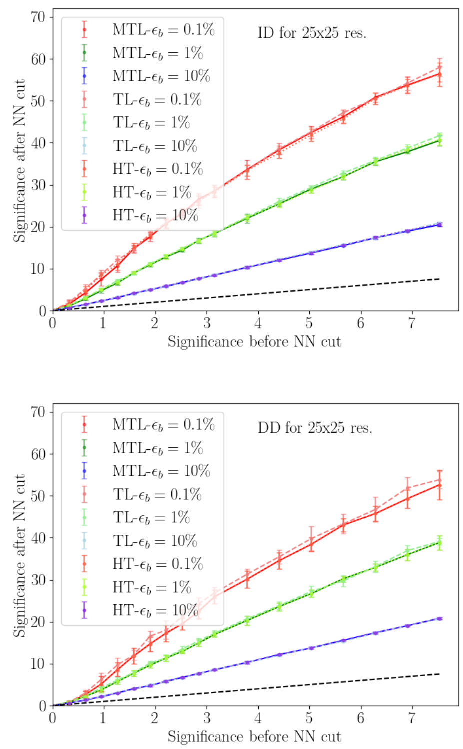

FIG. 3. Meta-transfer learning and hard tasks results for the ID10 and DD10 scenario, 25x25 resolution.

and hard tasks inclusion is inspired by [4], and the implementation of meta-transfer learning followed [1] . For hard tasks implementation, I chose 6 benchmarks with the lowest accuracies, and trained the model for one further epoch for each of the benchmarks.

The results are presented in Fig. **3**. I found that meta-transfer learning didn't result in a notable improvement over transfer learning. This is because transfer learning already achieved nearly the mathematical limit of neural network performance. The inclusion of hard tasks also didn't help with the performance. However, it is worth noting that in some cases, with a good choice of kernel, meta-transfer learning can perform significantly better, such as in the original paper [1].

IV. EXPLORATION

One of the problems with the CWoLa method is the dependence of certain variables on the dijet invariant mass. This can cause the neural network to learn the difference between the signal and sideband regions rather than learning the true difference between signal and background. Therefore, a suitable variable needs to be as loosely correlated to the dijet invariant mass as possible, while maintaining a difference between background and signal distribution. I found out that in the simulated signal and background data, the two jets' masses and Les Houches angularities (a measurement of the jet's angular distribution) were variables that satisfy the above conditions. The jet mass' and Les Houches angularity's independence from dijet invariant mass is demonstrated in Fig. **4** and Fig. **5**, where the histogram distribution is similar between background in the SR and SB region but is different between background and signal in the SR region. These variables are also previously used by [5] and [6] to distinguish between different jet structures using CWoLA.

To incorporate these new variables as inputs into the neural network, a plausible implementation is to create a parallel multi perceptron neural network that takes the additional variables as input and provides a number as output. This number can be multiplied with the number originally produced by the processing of jet images to provide a prediction. I have explored ways to combine the two features directly into the jet images' neural network, but these tend to result in the neural network learning to distinguish between SR and SB regions rather than between signal and background.

The results are presented in Fig. **5**. For CWoLA, the additional parameters assisted the neural network in the low significance regime, which might be attributed to the fact that the neural network requires a decent amount of signal in the mixed data to pick up the image features.

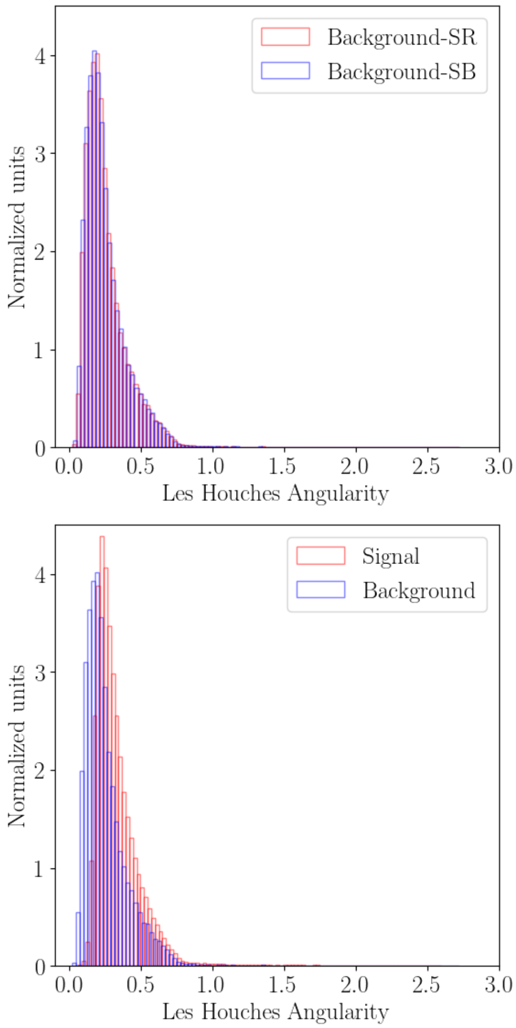

FIG. 4. Normalized distribution of first jet's Les Houches angularity

However, in the high significance regime, the neural network performed worse with the introduction of additional parameters. Compared to the exponential behavior seen before in the graph with only jet images, the mixed data neural network exhibited a linear trend, which suggested that Les Houches angularity and jet mass played the primary role in informing the neural network's classification. Transfer learning with additional variables also produced inferior results compared to only jet images. This behavior is highly puzzling, and future work will tackle this problem.

V. ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Cheng-Wei Chiang, Zong-En Chen, and Feng-Yang Hsieh for their unwavering support during my time in the lab, from providing directions for my project to reviewing

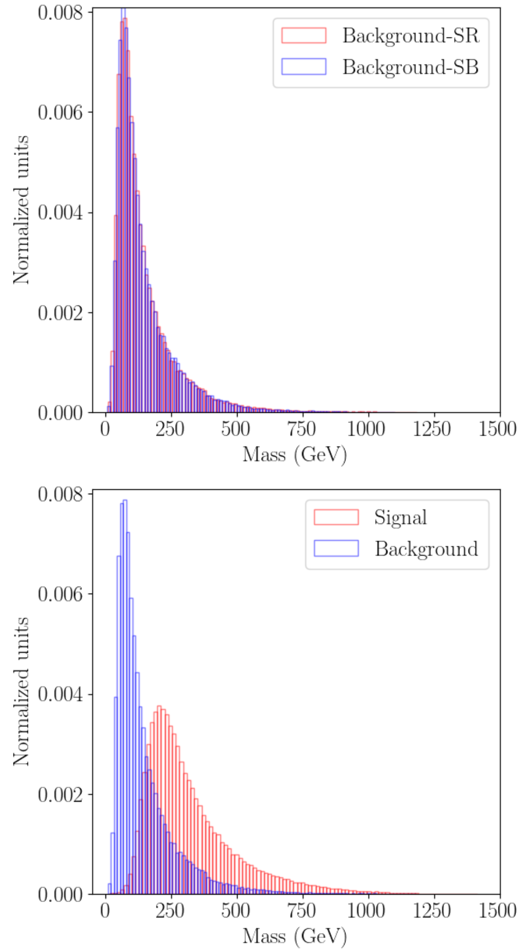

FIG. 5. Normalized distribution of first jet's mass

codes and offering feedback for the presentation and report. I also want to say thank you to Professor Cheng Chin, professors at NTU, and members of the OIA who have organized this wonderful program. Finally, this summer at Taiwan would not be one of my best summers ever without the assistance and encouragement from my family and friends, including other fellows, my labmates, and NTU buddies.

[1] H. Beauchesne, Z.-E. Chen, and C.-W. Chiang, Improving the performance of weak supervision searches using transfer and meta-learning, Q. J. Mech. Appl. Math. **02**, 138 (2024).

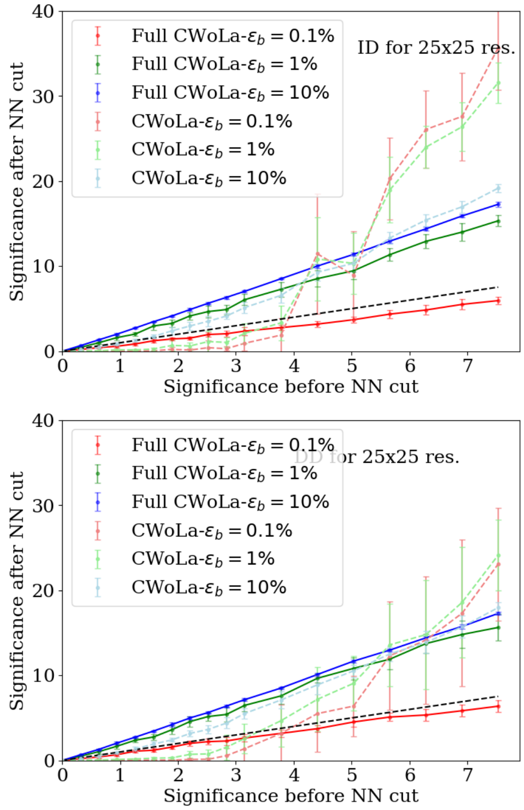

FIG. 6. CWoLA results for the ID10 and DD10 scenario, 25x25 resolution. Full CWoLA refers to the model with additional parameters, whereas CWoLa refers to the original model training on only jet images.

[2] ATLAS-Collaboration, Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, Phys. Lett. B **716**, 1 (2012).

[3] E. Metodiev, B. Nachman, and J. Thaler, Classification without labels: learning from mixed samples in high energy physics, JHEP **2017**, 174.

[4] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, Meta-transfer learning for few-shot learning, IEEE Computer Society , 403 (2019).

[5] CMS-collaboration., A. Tumasyan, and W. A. et al, Study of quark and gluon jet substructure in z+jet and dijet events from pp collisions, JHEP **2022**, 188.

[6] J. Collins, K. Howe, and B. Nachman, Anomaly detection for resonant new physics with machine learning, Phys. Rev. Lett. **121** (2018).