

RECEIVED: July 12, 2018

REVISED: February 8, 2019

ACCEPTED: February 16, 2019

PUBLISHED: February 20, 2019

# Infrared safety of a neural-net top tagging algorithm

Suyong Choi,<sup>a</sup> Seung J. Lee<sup>a,b</sup> and Maxim Perelstein<sup>c</sup>

<sup>a</sup>*Department of Physics, Korea University,  
Seoul 02841, Republic of Korea*

<sup>b</sup>*School of Physics, Korea Institute for Advanced Study,  
Seoul 130-722, Korea*

<sup>c</sup>*Laboratory for Elementary Particle Physics, Cornell University,  
Ithaca, NY 14853, U.S.A.*

*E-mail:* [suyong@korea.ac.kr](mailto:suyong@korea.ac.kr), [sjjlee@korea.ac.kr](mailto:sjjlee@korea.ac.kr), [mp325@cornell.edu](mailto:mp325@cornell.edu)

**ABSTRACT:** Neural network-based algorithms provide a promising approach to jet classification problems, such as boosted top jet tagging. To date, NN-based top taggers demonstrated excellent performance in Monte Carlo studies. In this paper, we construct a top-jet tagger based on a Convolutional Neural Network (CNN), and apply it to parton-level boosted top samples, with and without an additional gluon in the final state. We show that the jet observable defined by the CNN obeys the canonical definition of infrared safety: it is unaffected by the presence of the extra gluon, as long as it is soft or collinear with one of the quarks. Our results indicate that the CNN tagger is robust with respect to possible mis-modeling of soft and collinear final-state radiation by Monte Carlo generators.

**KEYWORDS:** Jets, QCD Phenomenology

**ARXIV EPRINT:** [1806.01263](https://arxiv.org/abs/1806.01263)

---

## Contents

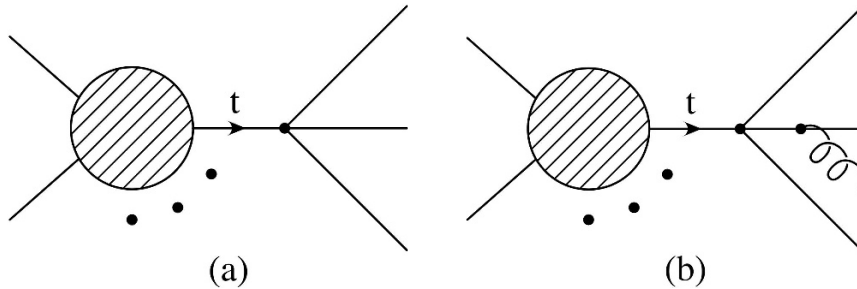
<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Neural net tagger and event samples</b>	<b>3</b>
<b>3</b>	<b>Results</b>	<b>6</b>
<b>4</b>	<b>Discussion</b>	<b>10</b>

---

## 1 Introduction

Events at the Large Hadron Collider (LHC) contain large numbers of jets. The jets can be classified into four types, according to their origin: (i) light-quark jets, initiated by  $u, d, s$  or  $c$  quarks; (ii) gluon jets; (iii)  $b$ -quark initiated jets; and (iv) jets created by a hadronic decay of a highly boosted massive object, such as a  $W/Z$  boson, Higgs, or top quark. In the latter case, hadronic showers created by each of the partons overlap, and standard jet reconstruction algorithms recognize them as a single merged jet. A *jet classification algorithm*, or “tagger”, attempts to reconstruct the origin of each individual jet, based on the information accessible to the experiment, i.e. detector-level data. Recently, there has been strong interest in applying modern machine-learning techniques, such as Neural Networks (NNs), to the jet classification problem. This is motivated as follows. The pattern of energy deposits in individual hadron calorimeter (HCAL) cells can be thought of as a two-dimensional image of the jet. Jets of each type have a characteristic shower history, resulting in differences in spatial distribution of energy inside the jet, often called “jet substructure”. The jet classification problem is thus mapped onto a 2D image recognition problem [1]. Application of NNs to image recognition is a well-developed field of computer science. Advanced NN-based image recognition techniques have been applied to jet classification problems in Monte Carlo (MC) studies, with highly promising results [2–7]. (For related applications of machine-learning techniques, see refs. [8–19]; for a recent review, see [20]). For example, NN-based top taggers have been shown to significantly outperform traditional top-tagging algorithms currently in use by the LHC experiments.

Can NN-based taggers trained on MC samples be used in real data analysis? The answer hinges on whether the features of jet substructure that are identified by the NN as important for classification are in fact accurately modeled by the MC generator. This is a non-trivial issue. Parton showering cannot be described by fixed-order perturbation theory, since soft and collinear parton splittings suffer from infrared/collinear (IRC) singularities. As a result, MC predictions of energy distribution within jets, in particular on small angular scales, suffer from significant (and poorly quantified) theoretical uncertainties. At the same time, unlike traditional taggers, the highly non-linear, multi-variable nature of the NN



**Figure 1.** Parton-level top decays, with and without a final-state gluon respectively.

tagger output makes it very difficult to identify the specific features in the jet substructure that the NN focuses on, let alone assess their robustness in the simulation. To date, this issue has been addressed by cross-comparisons of NN taggers trained on samples produced by different MC generators, which employ different algorithms to model parton showers (see e.g. refs. [2, 5]). While the results seem to indicate that the NN output is robust, a deeper understanding of this issue is clearly desirable to put this approach to jet tagging on a firm foundation.<sup>1</sup>

Traditionally, observables in jet physics are thought to be robust with respect to uncertainties in parton shower modeling if they satisfy the requirement of *Infrared/Collinear (IRC) Safety*. The notion of IRC safety applies to parton-level events. An observable  $\mathcal{O}$  is IRC safe if a soft or collinear splitting of one of the partons leaves  $\mathcal{O}$  unchanged:

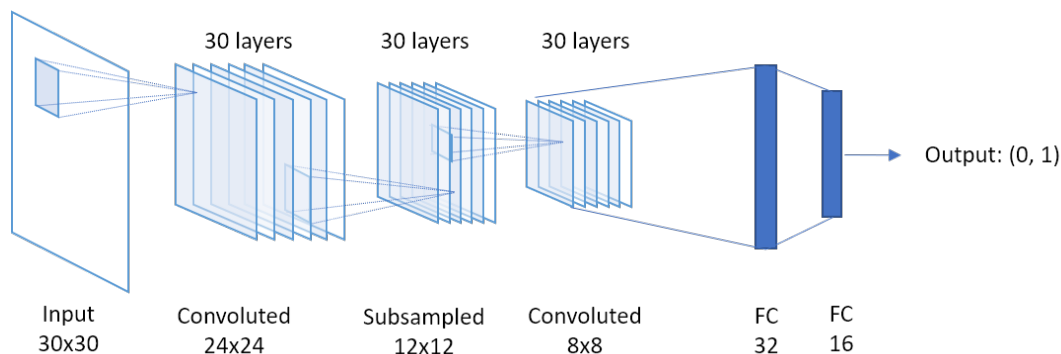
$$\mathcal{O}_n(p_1, \dots, p_i, p_{i+1}, \dots, p_n) \rightarrow \mathcal{O}_{n-1}(p_1, \dots, p_i + p_{i+1}, \dots, p_n) \quad (1.1)$$

whenever  $p_{i+1}$  becomes soft or collinear with  $p_i$ . For example, consider the two events shown in figure 1. In the limit when gluon in figure 1(b) becomes either soft ( $p_{T,g} \rightarrow 0$ ) or collinear with one of the quarks ( $p_g \cdot p_i \rightarrow 0$ ), the value of  $\mathcal{O}$  evaluated on the final state (b) should approach its value evaluated on the final state (a).

It’s worth noting that IRC safety is not the necessary condition for calculability of physical observables. Sudakov safe observable [22, 23] is a notable example which is IRC unsafe, but calculable if all-orders effects are resummed. The numerical analysis of this paper cannot distinguish between IRC-safe and “unsafe but calculable”, Sudakov-safe observables, because Sudakov-suppressed regions of phase space are not probed. While we will use the term “IRC safety” throughout the paper for brevity, this caveat should be kept in mind.

NN tagger is an observable that maps the matrix of energy deposits in individual HCAL cells onto a number between 0 and 1, the “topness” of the jet. The goal of this paper is to check whether this observable is IRC safe. We perform this test in the particular context of a Convolutional Neural Network (CNN) top tagger. The CNN is first trained on *particle-level* (showered and hadronized) MC samples of boosted top jets and “QCD”

<sup>1</sup>An alternative would be to avoid the use of MC generators altogether by training directly on real data. This would require one to identify training sets, tagged by an object external to the jet, in the data. For recent interesting work in that direction, see refs. [11, 21].



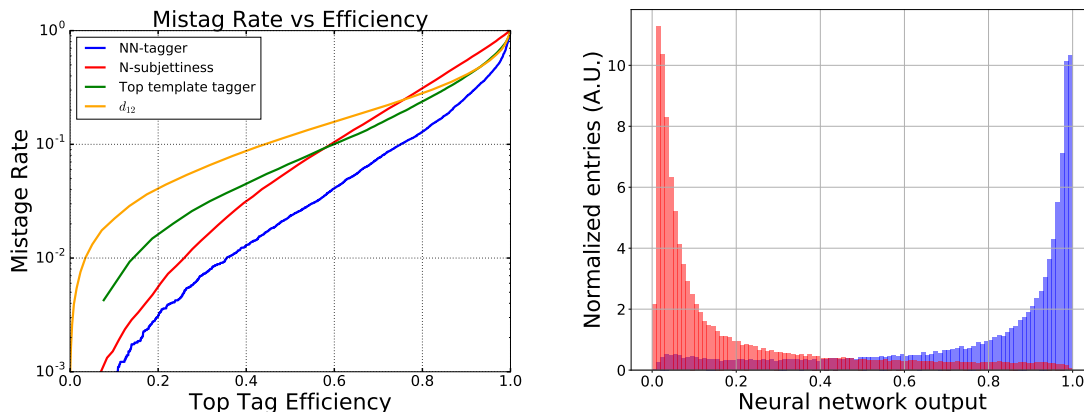
**Figure 2.** Convolutional Neural Network (CNN) architecture used for boosted top-jet tagging.

(light quark/gluon) jets. We then apply this CNN to *parton-level* hadronic top events. This defines a parton-level observable, to which the above canonical definition of IRC safety can be applied. We study the behavior of this observable as a function of the gluon momentum and collinearity in the  $(t \rightarrow 3q) + g$  sample shown figure 1(b). Our numerical results strongly support the hypothesis that the CNN output is IRC-safe. The training process appears to result in a network that largely disregards small-scale angular features in the energy distribution inside the jet, making the CNN tagger robust with respect to modeling such small-scale features in MC generators. Such robustness is a necessary pre-condition for practical applicability of MC-trained NN taggers, and it is highly reassuring that it is satisfied.

The rest of the paper is organized as follows. In section 2, we discuss the architecture of the CNN tagger, its training and performance on particle-level MC samples. We also describe the parton-level “merged” and “unmerged” samples used for numerical tests of IRC safety of the tagger. The main results of the analysis are presented in section 3, which contains the evidence to support our claim that the CNN observable is IRC-safe according to the canonical definition, eq. (1.1). Discussion of the results and conclusions are contained in section 4.

## 2 Neural net tagger and event samples

The top tagger used in this study consists of a Convolutional Neural Network (CNN), which has proven to be one of the best performers in problems of pixelated image recognition. CNN architecture is known to produce robust identification of translationally invariant features of *a priori* unknown size. In our case, the feature of interest is subjects, which clearly play an important role in top tagging. The network architecture is schematically shown in figure 2. We used the `mxnet` software package for implementation of the CNN [24] on a NVIDIA Geforce GTX 1080 GPU. The input layer of the CNN is the Hadronic Calorimeter (HCAL), modeled as a set of  $30 \times 30$  square pixels of size  $(\Delta\phi, \Delta\eta) = (0.1, 0.1)$ . The pixels are populated by normalized energy deposited in each bin by a jet, preprocessed according to the procedure used in ref. [2]. Preprocessing places the center of the jet at the center of the image, and rotates the jet so that the principal axis always has the same orientation. In



**Figure 3.** Left panel: performance of CNN, compared with other variables for boosted top-jet tagging: Template Tagger [25–27] and  $N$ -subjettiness [28, 29]. Right panel: neural network output distributions for the signal and background.

this way, overall translational and rotational symmetries of the jet are factored out and do not need to be learned in the training process.<sup>2</sup> The next layer consists of thirty “filters”, which are convoluted with the input image according to

$$o_\ell(i, j) = \sum_{x=1}^7 \sum_{y=1}^7 I(i+x, j+y) w_\ell(x, y) \quad (i, j = 1, \dots, 24, \ell = 1, \dots, 30), \quad (2.1)$$

where  $I$  is the input image,  $w_\ell$  are the filters, and  $o_\ell$  is the output of the convolution operation. We use ReLU activation function. The individual weights of the filters,  $w_\ell(x, y)$  are determined during training of the CNN, using back-propagation methods. Each filter is to learn some distinguishing features that separate signal from backgrounds. The outputs of these layers are subsampled and convoluted further with different set of filters, as shown in figure 2. Ultimately, the final fully connected layer produces a single output, the “topness” of the jet  $Y \in [0, 1]$ , with  $Y = 1$  corresponding to a boosted top jet and  $Y = 0$  corresponding to a QCD jet.

To train the network, we use the MC samples of particle-level top and QCD jets that were previously used in ref. [2]. The samples were generated using MadGraph 5 [30] at parton level ( $t\bar{t}$  and  $q\bar{q}$  final states), following by showering and hadronization with Pythia 8 [31]. We use a sample of jets reconstructed using anti- $k_T$  algorithm with a large jet cone of  $R = 1.0$ , with  $p_T$  in 800–900 GeV range. We further require that the jet mass ( $m_J$ ) be in the range of 130–210 GeV. A majority of top-jets would fall in this mass range, while most QCD jets would be rejected by the  $m_J$  requirement. The top and QCD jets passing these basic cuts are preprocessed as described in [2], and provided as inputs to the CNN. For training, weights were initialized using Xavier initialization [32], and Adam hyperparameter optimization [33] with learning rate of 0.05 and dropout regularization rate

<sup>2</sup>As an alternative, one can use  $p_T$ , instead of energy, as the input observable, motivated by Lorentz invariance [3]. We explored this alternative and did not find a significant difference in training time or performance.

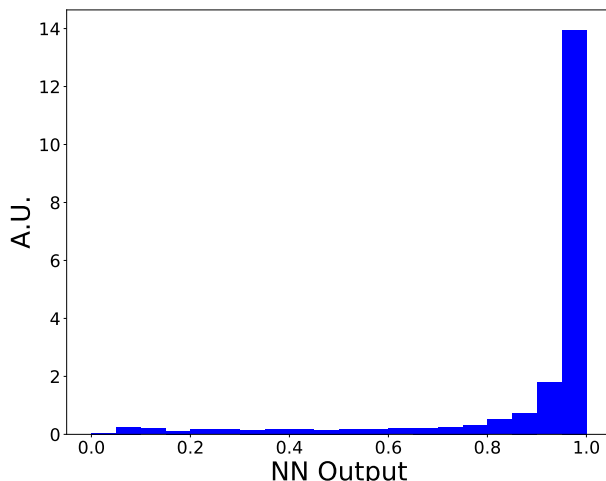
of 0.0001 was used. 100 epochs through the training samples were made using minibatches of 1000 events. After training, the CNN performance was evaluated using the test samples, generated and preprocessed in the same way as the training samples. The results are shown in figure 3. The results are not very sensitive on the choices of hyperparameters during training. The CNN output provides a clear separation between the two types of jets. For interesting top tagging efficiencies, the mistag rate is reduced by a factor of almost 2 compared to traditional observables, providing further improvement upon the two hidden layer multi-layer perceptron type deep neural network used in [2].

As explained in the Introduction, the notion of IRC safety applies to observables defined on parton-level events. The observable we want to study is the output of the CNN, the “topness”  $Y \in [0, 1]$ . The CNN maps a set of energy deposits in individual pixels into this observable:  $I(i, j) \rightarrow Y$ . A parton-level event of the type shown in figure 1 is trivially mapped into  $I(i, j)$  by identifying each parton’s location in the  $(\eta, \phi)$  space, and assigning the value of that parton’s energy to the corresponding HCAL cell. This defines the action of the CNN on parton-level events, which can be thought of as a map

$$\mathcal{O} : \{p_i, i = 1 \dots N_p\} \rightarrow Y, \quad (2.2)$$

where  $p_i$  is the parton 4-momenta, and  $N_p$  is the number of partons in the event. We would like to study whether the IRC safety criterion, eq. (1.1), applies to this map. While  $\mathcal{O}$  is a completely well-defined function, it is horrendously complicated and highly non-linear, making an analytic study of its limits impractical. Instead, we will check the IRC safety criterion numerically. To this end, we used `MadGraph` [30] to generate a parton-level sample of hadronically decaying top quarks, with an additional gluon in the final state, as in figure 1(b). (To avoid unnecessary complexity, we simulate a process with no other colored particles in the final state.) In this simulation, cuts on the gluon momentum and its separation from each quark must be imposed to avoid infinities associated with soft and collinear singularities. Since we are primarily interested in precisely the gluons in the soft and collinear regions, the cuts we impose are very low:  $p_T \geq 5 \text{ GeV}$ ,  $\Delta R_{qg} \geq 0.05$ . One may question whether a fixed-order simulation correctly approximates the cross section for such low values of  $p_T$  and  $\Delta R_{qg}$ . For our purposes, however, this question is irrelevant. We want to study how the CNN response is affected by the presence of a soft or collinear gluon, and the purpose of the simulation is simply to provide a sample of such events; we do not use any information about their overall cross section or phase-space distribution.

To ensure that the CNN is applied in the same regime where it was trained, we compute the “jet  $p_T$ ” (the sum of the four parton  $p_T$ ’s) and the “jet invariant mass” (the total invariant mass of the four partons) for each event, and apply the same cuts as in the training sample,  $p_T \in [800\text{--}900] \text{ GeV}$ ,  $m_J \in [130\text{--}210] \text{ GeV}$ . The sample constructed in this way is referred to as the *unmerged sample*. We construct the *merged sample* by taking each event in the unmerged sample, identifying the quark closest to the gluon (in terms of  $\Delta R_{ij}$  separation), and replacing that quark and the gluon with a single parton with 4-momentum equal to the sum of the two. Applying the CNN map to the unmerged and merged samples corresponds to evaluating the left-hand side and the right-hand side of



**Figure 4.** Neural network output distribution on parton-level top sample.

eq. (1.1), respectively. Checking the IRC safety criterion then amounts to comparing the CNN outputs on these two samples.

### 3 Results

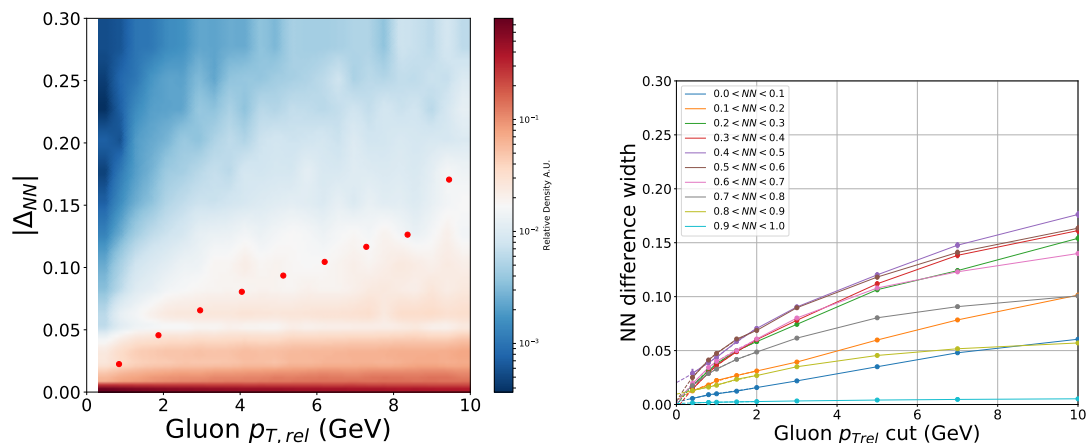
Training the CNN on particle-level top and QCD samples and applying it to the parton-level top sample produces the output distribution shown in figure 4. Clearly, the network predominantly still perceives such events as top-like. The fact that the NN did not need to be retrained when a switch from particle-level to parton-level input was made already provides some evidence that the observable defined by the CNN is likely IRC-safe. In the rest of this section, we will attempt to establish the IRC safety more directly, by comparing CNN outputs on merged and unmerged samples as explained above.

To gauge the impact of soft/collinear gluon radiation, we compute the difference  $\Delta_{\text{NN}}$  between the CNN output from an event in the unmerged sample and the corresponding event in the merged sample. A convenient measure of soft/collinear kinematics of the gluon is provided by its “relative  $p_T$ ”, defined by

$$p_T^g = \left| \mathbf{p}_g - \frac{\mathbf{p}_g \cdot \mathbf{p}_q}{|\mathbf{p}_q|^2} \mathbf{p}_q \right|, \tag{3.1}$$

where  $\mathbf{p}_q$  is the 3-momentum of the quark nearest (in terms of  $\Delta R_{qg}$  separation) to the gluon. Physically,  $p_T^g$  is the component of the gluon 3-momentum transverse to the nearest quark, and it vanishes in both soft and collinear limits. If the CNN observable is IRC safe, we expect  $\Delta_{\text{NN}}$  to go to zero in the limit of vanishing relative  $p_T$ . The distribution of  $|\Delta_{\text{NN}}|$  and  $p_T^g$  values in our event sample is shown in the left panel of figure 5. For most events,  $|\Delta_{\text{NN}}|$  is small, which is reassuring: adding a soft gluon does not lead to a dramatic change in the CNN output. There is, however, a tail of events where the change is significant. To better characterize this tail, we bin the data in relative  $p_T$  and calculate the *width* of the  $|\Delta_{\text{NN}}|$  distribution in each bin. The width  $|\Delta_{\text{NN}}|_{90}$  for each bin is defined by requiring that





**Figure 5.** Left panel: difference in CNN output between merged and unmerged events,  $|\Delta_{NN}|$ , as a function of the gluon transverse momentum relative to its nearest quark,  $p_{T,rel}^g$ . Red dots show the width of the  $|\Delta_{NN}|$  distribution. Background colors indicate the relative density of events for given  $p_{T,rel}^g$ . Right panel:  $|\Delta_{NN}|$  width as a function of  $p_{T,rel}^g$ , shown separately for 10 NN output bins. The lines indicate an interpolating curve (third-order polynomial) fit to the data in each NN output bin.

90% of the events in that bin have  $|\Delta_{NN}| \leq |\Delta_{NN}|_{90}$ . The values of  $|\Delta_{NN}|_{90}$  are plotted as red dots in figure 5. The data exhibits a clear correlation between decreasing relative  $p_T$  and decreasing width, indicative of IRC safety. In fact, the data is consistent with the hypothesis that  $|\Delta_{NN}|_{90} \rightarrow 0$  in the limit of  $p_{T,rel}^g \rightarrow 0$ .

In the right panel of figure 5, the data is further subdivided into 10 bins according to the NN output evaluated on the merged sample, and dependence of the width on relative  $p_T$  is shown separately for each bin.<sup>3</sup> For events in the last bin,  $0.9 \leq Y \leq 1$ , emission of an extra gluon has almost no effect even if it has a relatively large relative  $p_T$ . This is presumably due to the fact that  $Y$  is already close to the upper boundary. The events in this bin are therefore consistent with the IRC safety hypothesis, but do not show much variation as relative  $p_T$  is varied. On the other hand, events in all other  $Y$  bins show a very clear convergence between the output values with and without the extra gluon in the  $p_{T,rel}^g \rightarrow 0$  limit.

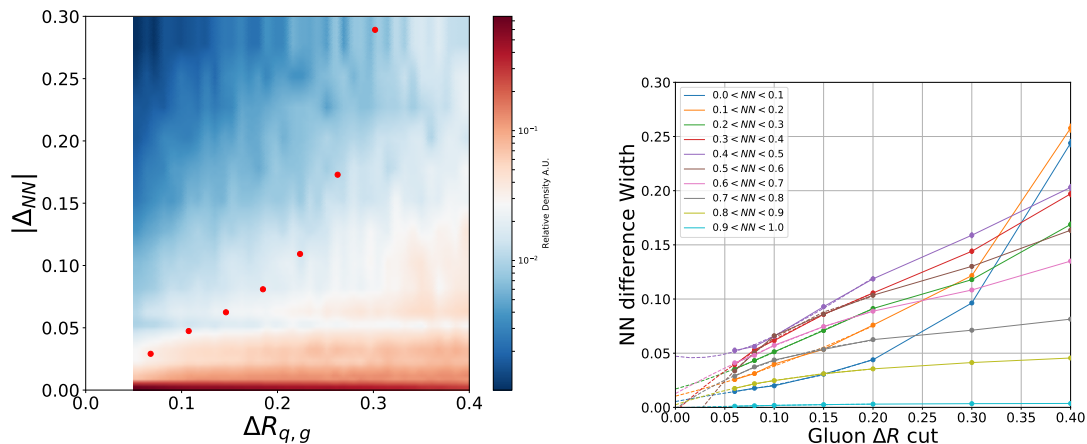
The relative  $p_T$  observable goes to zero in both soft and collinear limits. It is interesting to probe the convergence of the CNN output in each of these limits separately. To this end, we study two observables. The first one is the angular separation between the gluon and the nearest quark,  $\Delta R_{qg}$ , which goes to zero in the collinear limit, but not the soft limit. The second one is the “longitudinal momentum ratio”, defined by

$$p_{L, \text{rat}} = \frac{\mathbf{p}_g \cdot \mathbf{p}_q}{\mathbf{p}_q \cdot \mathbf{p}_q}, \quad (3.2)$$

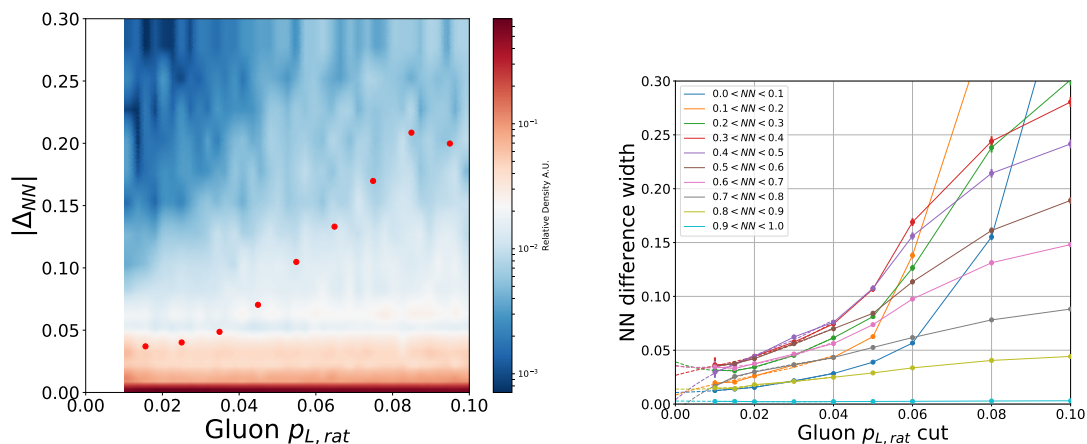
where  $\mathbf{p}_q$  is the 3-momentum of the quark nearest (in terms of  $\Delta R_{qg}$  separation) to the

<sup>3</sup>In some bins, the distribution of  $\Delta_{NN}$  is sharply asymmetric around zero, mainly because  $Y$  is restricted to lie between 0 and 1 by construction. To account for this, in the right panel of figure 5 we do not take the absolute value of  $\Delta_{NN}$ , but instead define the “NN difference width” as the width of the minimal interval containing 90% of events, not necessarily centered at 0.





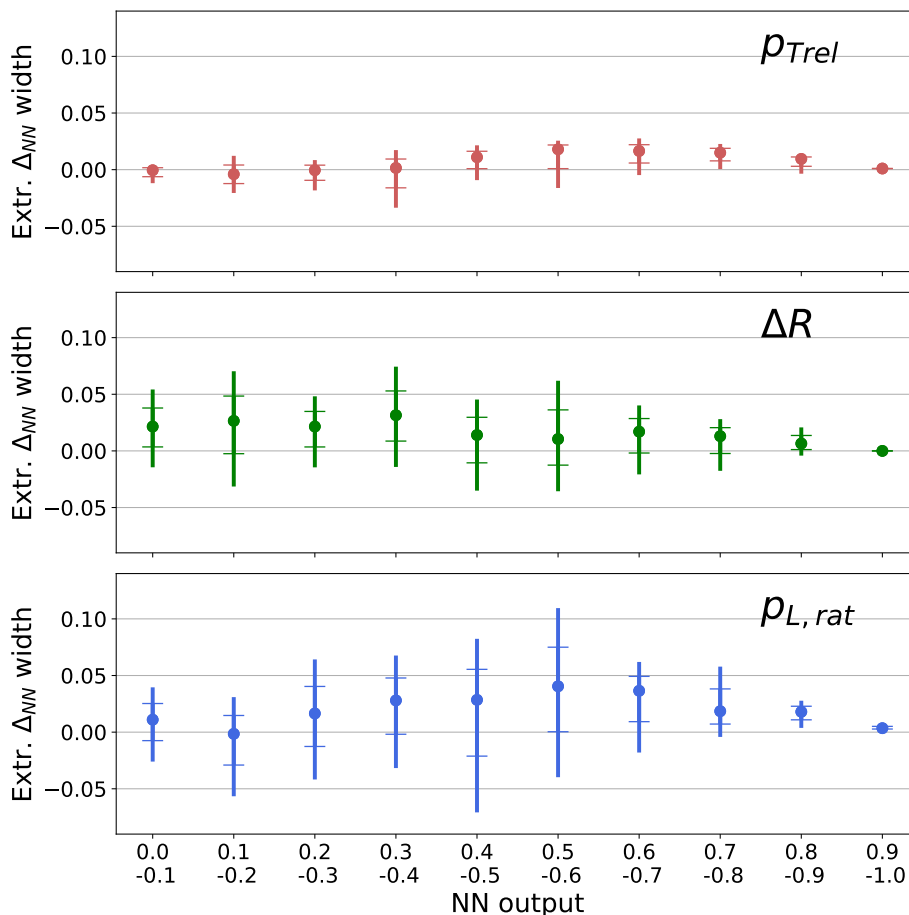
**Figure 6.** Left panel: difference in CNN output between merged and unmerged events,  $|\Delta_{NN}|$ , as a function of the gluon’s angular separation from its nearest quark,  $\Delta R_{qg}$ . Red dots show the width of the  $|\Delta_{NN}|$  distribution. Background colors indicate the relative density of events for given  $\Delta R_{qg}$ . Right panel:  $|\Delta_{NN}|$  width as a function of  $\Delta R_{qg}$ , binned in 10 NN output intervals. The lines indicate an interpolating curve (third-order polynomial) fit to the data in each NN output bin.



**Figure 7.** Left panel: difference in CNN output between merged and unmerged events,  $|\Delta_{NN}|$ , as a function of the longitudinal momentum ratio defined in eq. (3.2). Red dots show the width of the  $|\Delta_{NN}|$  distribution. Background colors indicate the relative density of events for given longitudinal momentum ratio. Right panel:  $|\Delta_{NN}|$  width as a function of longitudinal momentum ratio, binned in 10 NN output intervals. The lines indicate an interpolating curve (third-order polynomial) fit to the data in each NN output bin.

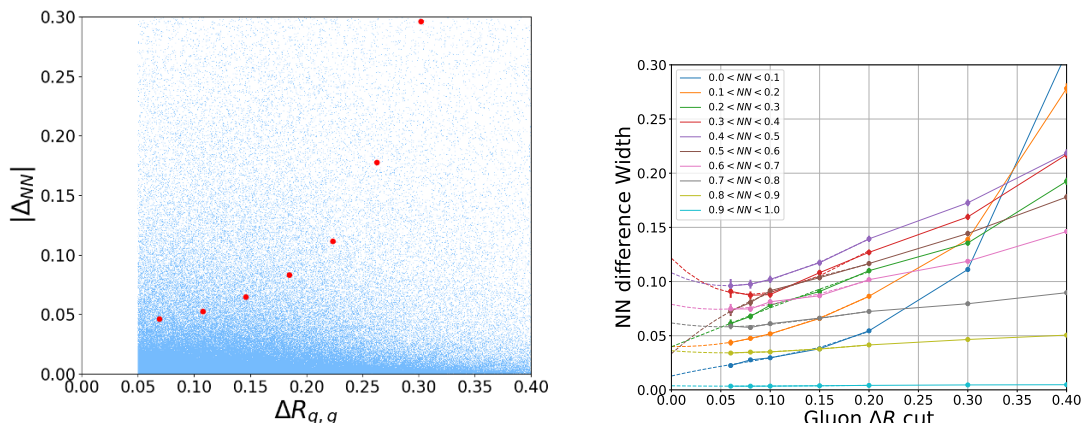
gluon. This observable vanishes when the gluon is soft, but not when it is collinear with one of the quarks. The difference in CNN outputs for merged and unmerged samples as a function of these two observables is shown in figures 6 and 7. We conclude that the convergence of the outputs holds separately in both soft and collinear limits.

The right panels of figures 5, 6, and 7 include interpolating functions obtained by fitting the data in each NN output bin with a third-order polynomial. Expected values of  $|\Delta_{NN}|$  in the exact IR(C) limit can be obtained by extrapolating these functions. Let us define



**Figure 8.** Extrapolated values of  $\Delta(0)$ , including 68% and 95% confidence intervals. Training and extrapolation uncertainties are combined in quadrature.

$\Delta(x)$  to denote the width of the  $|\Delta_{\text{NN}}|$  distribution in the bin centered at  $x$ , where  $x = p_T^g$ ,  $\Delta R_{qg}$ , or  $p_{L, \text{rat}}$ . In all three cases,  $x = 0$  corresponds to the exact IR(C) limit, so that  $\Delta(0) = 0$  would indicate exact IR(C) safety. It is clear from figures 5–7 that  $\Delta(0) \leq 0.02$  in most cases, and  $< 0.05$  in all cases. Further, since these values are obtained numerically, they are inherently uncertain. An important uncertainty is introduced by the NN tagger training process, which converges to slightly different weight configurations depending on the random weight initialization at the start of training. To estimate this uncertainty, we trained an ensemble of 20 identical NN taggers, with the only difference being the random initial weights. We then repeated the above analysis for each NN in the ensemble, and used the spread in the values of  $\Delta(0)$  measured within this ensemble as an estimate of the training uncertainty. Further, we estimated the uncertainty due to extrapolation from the measured values of  $\Delta$  at finite  $x$  to  $x = 0$ , and combined it with the training uncertainty in quadratures to estimate the total uncertainty. The measured values of  $\Delta(0)$  including this uncertainty are shown in figure 8. Our data is consistent with the hypothesis that  $\Delta(0) = 0$ . (The systematic preference for positive offset is due to the definition of  $\Delta$ , which is defined to be non-negative at all points where it is measured;  $\Delta(0) < 0$  occurs only as an artifact of extrapolation.)



**Figure 9.** Same distributions as in figure 6. Events where the gluon and its nearest quark are in the same pixel have been removed from the sample.

Our analysis shows a clear trend towards convergence of the NN output with and without an extra gluon, in the limit when the extra gluon is soft or collinear, and extrapolation to the exact soft/collinear limit is consistent with exact IR(C) safety. The analysis leaves the room open for  $\mathcal{O}(\text{few } \%)$  deviations from exact IR(C) safety, due to limitations inherent in the numerical approach to this issue. However, we note that in practice, such small deviations will not change the outcome of the tagging procedure for a vast majority of jets. For example, if the tagging threshold is chosen to be  $Y = 0.55$ , we find that only 0.7% of events in the top sample lie within  $\Delta Y = \pm 0.02$  of the threshold, while 1.8% lie within  $\Delta Y = \pm 0.05$ , meaning that the classification of 98 – 99% of top jets would be unaffected.

By construction, events in which the extra gluon lands in the same pixel as its nearest quark will have  $\Delta_{NN} = 0$ . This feature makes the CNN observable automatically IRC-safe in the limit of small  $\Delta R$ . Is the observed IRC safety in this limit due entirely to this feature? To address this question, we repeated the analysis on a sample in which events where the gluon and a quark are in the same pixel have been removed. The result is shown in figure 9. The trend towards the CNN outputs converging in the  $\Delta R \rightarrow 0$  limit persists in this sample. This indicates that the CNN output for a sample with an extra gluon converges smoothly as the gluon approaches its nearest quark, even if they do not land in the same cell. Such convergence is a non-trivial feature of CNN’s treatment of energy patterns, and not just a trivial consequence of finite cell size.

## 4 Discussion

Starting with ref. [2], many studies have demonstrated the efficacy of Neural Networks for boosted top jet tagging, at the level of Monte Carlo (MC) simulations. All studies to date have trained and evaluated NN top taggers using particle-level MC samples of top and QCD jets. In this paper, a Convolutional Neural Network (CNN) top jet tagger was constructed. While particle-level MC samples were used in training as usual, we then applied the resulting tagger to a sample of *parton-level* top events with and without an

additional gluon in the final state, as shown in figure 1. We showed that this observable obeys the Infrared Safety criterion: the output of the CNN applied to an event with an extra gluon approaches its output on the same event without an extra gluon, in the limit when the extra gluon becomes soft or collinear with one of the quarks.

Our analysis does not constitute a complete proof of IRC safety of the CNN output. The reason is that we studied only one class of final states, those containing a high- $p_T$  top quark. In the language of eq. (1.1), our analysis demonstrates IRC safety for some configurations of final-state parton momenta  $p_i$ , but not for general  $p_i$ . We did not study the behavior of the CNN output on non-top events, mainly because there is a very broad range of possible momentum configurations, making a comprehensive numerical study impractical. Moreover, it is not immediately clear which of these configurations are most important for the top tagging problem. We hope to be able to address this issue in future work.

In spite of this limitation, the results of the analysis presented in this paper are highly reassuring. Certainly, the value of NN-based approach to top jet tagging would be in very serious doubt if an addition of a soft or collinear parton to the final state led to an order-one change in the NN output. We showed that this is not the case for the events with a genuine high- $p_T$  top quark in the final state. This result places the NN-based taggers on firmer foundation, and should provide encouragement for further development of this approach.

## Acknowledgments

The authors are grateful for conversations with Mihailo Backovic, Steven Durr, and Jesse Thaler. MP acknowledges the support of the U.S. National Science Foundation through grant PHY-1719877. SC was supported by National Research Foundation of Korea (NRF) grant NRF-2018R1A2B6005043. SL was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07049148), by the Korea government (MEST) (NRF-2015R1A2A1A15052408), and Samsung Science and Technology Foundation under Project Number SSTF-BA1601-07.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] J. Cogan, M. Kagan, E. Strauss and A. Schwartzman, *Jet-Images: Computer Vision Inspired Techniques for Jet Tagging*, *JHEP* **02** (2015) 118 [[arXiv:1407.5675](https://arxiv.org/abs/1407.5675)] [[INSPIRE](https://inspirehep.net/literature/1250000)].
- [2] L.G. Almeida, M. Backović, M. Cliche, S.J. Lee and M. Perelstein, *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, *JHEP* **07** (2015) 086 [[arXiv:1501.05968](https://arxiv.org/abs/1501.05968)] [[INSPIRE](https://inspirehep.net/literature/1250000)].
- [3] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *JHEP* **07** (2016) 069 [[arXiv:1511.05190](https://arxiv.org/abs/1511.05190)] [[INSPIRE](https://inspirehep.net/literature/1400000)].

- [4] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev. D* **93** (2016) 094034 [[arXiv:1603.09349](#)] [[INSPIRE](#)].
- [5] J. Barnard, E.N. Dawe, M.J. Dolan and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, *Phys. Rev. D* **95** (2017) 014018 [[arXiv:1609.00607](#)] [[INSPIRE](#)].
- [6] P.T. Komiske, E.M. Metodiev and M.D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, *JHEP* **01** (2017) 110 [[arXiv:1612.01551](#)] [[INSPIRE](#)].
- [7] G. Kasieczka, T. Plehn, M. Russell and T. Schell, *Deep-learning Top Taggers or The End of QCD?*, *JHEP* **05** (2017) 006 [[arXiv:1701.08784](#)] [[INSPIRE](#)].
- [8] G. Louppe, K. Cho, C. Becot and K. Cranmer, *QCD-Aware Recursive Neural Networks for Jet Physics*, *JHEP* **01** (2019) 057 [[arXiv:1702.00748](#)] [[INSPIRE](#)].
- [9] J. Pearkes, W. Fedorko, A. Lister and C. Gay, *Jet Constituents for Deep Neural Network Based Top Quark Tagging*, [arXiv:1704.02124](#) [[INSPIRE](#)].
- [10] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned Top Tagging with a Lorentz Layer*, *SciPost Phys.* **5** (2018) 028 [[arXiv:1707.08966](#)] [[INSPIRE](#)].
- [11] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, *JHEP* **10** (2017) 174 [[arXiv:1708.02949](#)] [[INSPIRE](#)].
- [12] K. Datta and A.J. Larkoski, *Novel Jet Observables from Machine Learning*, *JHEP* **03** (2018) 086 [[arXiv:1710.01305](#)] [[INSPIRE](#)].
- [13] T. Cheng, *Recursive Neural Networks in Quark/Gluon Tagging*, *Comput. Softw. Big Sci.* **2** (2018) 3 [[arXiv:1711.02633](#)] [[INSPIRE](#)].
- [14] S. Egan, W. Fedorko, A. Lister, J. Pearkes and C. Gay, *Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC*, [arXiv:1711.09059](#) [[INSPIRE](#)].
- [15] H. Lüo, M.-x. Luo, K. Wang, T. Xu and G. Zhu, *Quark jet versus gluon jet: deep neural networks with high-level features*, [arXiv:1712.03634](#) [[INSPIRE](#)].
- [16] S. Macaluso and D. Shih, *Pulling Out All the Tops with Computer Vision and Deep Learning*, *JHEP* **10** (2018) 121 [[arXiv:1803.00107](#)] [[INSPIRE](#)].
- [17] K. Fraser and M.D. Schwartz, *Jet Charge and Machine Learning*, *JHEP* **10** (2018) 093 [[arXiv:1803.08066](#)] [[INSPIRE](#)].
- [18] T. Roxlo and M. Reece, *Opening the black box of neural nets: case studies in stop/top discrimination*, [arXiv:1804.09278](#) [[INSPIRE](#)].
- [19] J.H. Collins, K. Howe and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, *Phys. Rev. Lett.* **121** (2018) 241803 [[arXiv:1805.02664](#)] [[INSPIRE](#)].
- [20] A.J. Larkoski, I. Moutl and B. Nachman, *Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning*, [arXiv:1709.04464](#) [[INSPIRE](#)].
- [21] P.T. Komiske, E.M. Metodiev, B. Nachman and M.D. Schwartz, *Learning to classify from impure samples with high-dimensional data*, *Phys. Rev. D* **98** (2018) 011502(R) [[arXiv:1801.10158](#)] [[INSPIRE](#)].
- [22] A.J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **09** (2013) 137 [[arXiv:1307.1699](#)] [[INSPIRE](#)].

- [23] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov Safety in Perturbative QCD*, *Phys. Rev. D* **91** (2015) 111501 [[arXiv:1502.01719](#)] [[INSPIRE](#)].
- [24] Apache MXNet, *Apache MXNet (Incubating). A flexible and efficient library for deep learning*, (2019) <http://mxnet.io/>.
- [25] L.G. Almeida, S.J. Lee, G. Perez, G. Sterman and I. Sung, *Template Overlap Method for Massive Jets*, *Phys. Rev. D* **82** (2010) 054034 [[arXiv:1006.2035](#)] [[INSPIRE](#)].
- [26] L.G. Almeida, O. Erdogan, J. Juknevich, S.J. Lee, G. Perez and G. Sterman, *Three-particle templates for a boosted Higgs boson*, *Phys. Rev. D* **85** (2012) 114046 [[arXiv:1112.1957](#)] [[INSPIRE](#)].
- [27] M. Backović and J. Juknevich, *TemplateTagger v1.0.0: A Template Matching Tool for Jet Substructure*, *Comput. Phys. Commun.* **185** (2014) 1322 [[arXiv:1212.2978](#)] [[INSPIRE](#)].
- [28] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with  $N$ -subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [29] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing  $N$ -subjettiness*, *JHEP* **02** (2012) 093 [[arXiv:1108.2701](#)] [[INSPIRE](#)].
- [30] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [[arXiv:1106.0522](#)] [[INSPIRE](#)].
- [31] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [32] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, in proceedings of the *Thirteenth International Conference on Artificial Intelligence and Statistics*, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010.
- [33] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].