

# Physics inspired feature engineering with Lorentz Boost Networks

M Erdmann, E Geiser, Y Rath and M Rieger

III. Physics Institute A, RWTH Aachen University, Otto-Blumenthal-Str., 52074 Aachen, Germany

E-mail: [yannik.rath@rwth-aachen.de](mailto:yannik.rath@rwth-aachen.de)

**Abstract.** We present a neural network architecture designed to autonomously create characteristic features of high energy physics collision events from basic four-vector information. It consists of two stages, the first of which we call the Lorentz Boost Network (LBN). The LBN creates composite particles and rest frames from the combination of final state particles, and then boosts said particles into their corresponding rest frames. From these boosted particles, characteristic features are created and used by the second network stage to solve a given physics problem. We apply our model to the task of separating top-quark pair associated Higgs boson events from a  $t\bar{t}$  background, and observe improved performance compared to using domain unspecific deep neural networks. We also investigate the learned combinations and boosts to gain insights into what the network is learning.

## 1. Introduction

A key part of the success of deep learning methods can be attributed to the development of dedicated architectures that exploit the underlying structure of a given task. For example, convolutional neural networks can extract characteristic features from image-like data by making use of translational invariance and local correlations.

To analyze collision events in particle physics, physicists often engineer characteristic variables that condense the relevant information of the event. While deep learning methods can be applied directly to raw data, combining both these specifically constructed and low-level variables often leads to improved performance [1, 2]. The aim of our model is to autonomize the creation of these types of characteristic variables using only raw particle information as input, by incorporating physics knowledge into our model architecture.

In high energy physics events, short lived particles such as top quarks or Higgs bosons can be created. These particles decay, possibly through a cascade of intermediate particles, into the final state particles measured in the detector. The initial and intermediate particles can be reconstructed from their decay products using energy-momentum conservation. The properties of these parent particles and their decay are prime candidates for characteristic variables of the event.

One complication is that the decay properties are distorted by the movement of the parent particle. However, they can be disentangled by performing a Lorentz boost into the parent particle's rest frame. Our model is built to exploit these properties by creating relevant combinations of particles and boosting them into appropriate rest frames.



In this paper, we first describe the architecture of our model. We present an example application of our network and compare its performance with that of standard deep neural networks. Finally, we investigate the physics-inspired parts of our architecture to gain insights into what the network is learning, and end with our conclusion.

## 2. Network architecture

Our network architecture is divided into two stages. In the first stage, the Lorentz Boost Network (LBN), characteristic features of the event are extracted from the four-vectors of the final state particles. In the second stage, these characteristic features are used to solve a specific physics task, such as the separation of signal and background events.

The complete model is depicted in fig. 1. First, linear combinations of the input four-vectors are built to construct composite particles and corresponding rest frames. The combinations are created using trainable weights. Each of the particles is then boosted into its corresponding rest frame. This transformation is performed by applying the boost matrix

$$\Lambda = \begin{bmatrix} \gamma & -\gamma\beta n_x & -\gamma\beta n_y & -\gamma\beta n_z \\ -\gamma\beta n_x & 1 + (\gamma - 1)n_x^2 & (\gamma - 1)n_x n_y & (\gamma - 1)n_x n_z \\ -\gamma\beta n_y & (\gamma - 1)n_y n_x & 1 + (\gamma - 1)n_y^2 & (\gamma - 1)n_y n_z \\ -\gamma\beta n_z & (\gamma - 1)n_z n_x & (\gamma - 1)n_z n_y & 1 + (\gamma - 1)n_z^2 \end{bmatrix} \quad (1)$$

to each particle, with the relativistic parameters  $\gamma = E/m$ ,  $\vec{\beta} = \vec{p}/E$ , and  $\vec{n} = \vec{\beta}/\beta$  of the rest frame. To calculate the transformation for many events efficiently, this matrix is rewritten as

$$\Lambda = I + (U \oplus \gamma) \odot ((U \oplus 1) \cdot \beta - U) \odot (e \cdot e^T) \quad (2)$$

with

$$U = \begin{bmatrix} -1^{1 \times 1} & 0^{1 \times 3} \\ 0^{3 \times 1} & -1^{3 \times 3} \end{bmatrix}, \quad e = \begin{bmatrix} 1^{1 \times 1} \\ -\vec{n}^{3 \times 1} \end{bmatrix}, \quad (3)$$

and the unit matrix  $I$ . The operators  $\oplus$  and  $\odot$  denote element-wise addition and multiplication. From the boosted particles, a generic set of features is extracted. These features can be separated into two different groups:

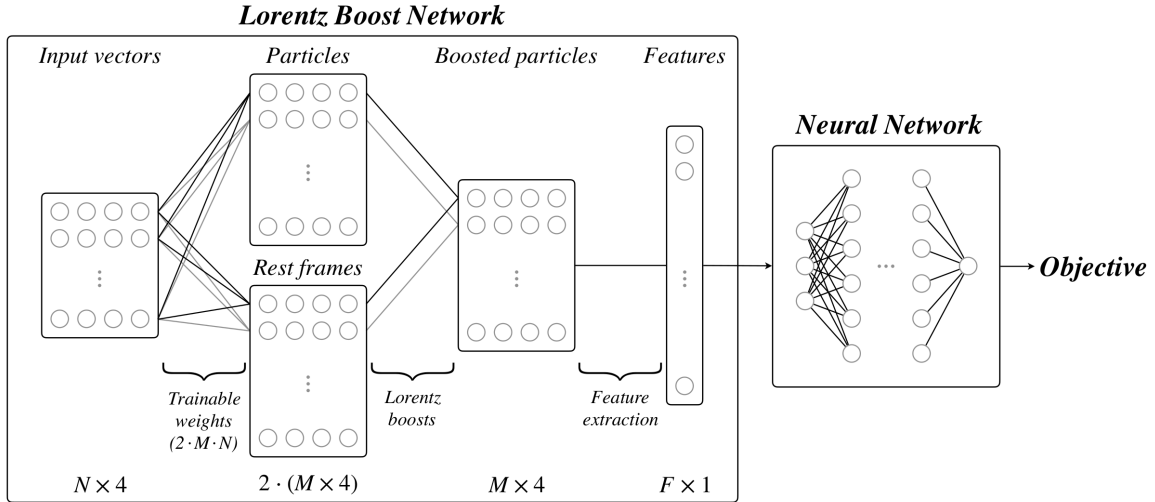
- (i) Features defined for a single boosted particle. These include the vector elements  $(E, p_x, p_y, p_z)$ , the mass, transverse and longitudinal momentum, pseudorapidity, and azimuthal angle.
- (ii) Pairwise features that combine information of two particles. These include the cosine of the spatial angle between the two particles, their distance in the  $\eta - \phi$  plane and their distance in Minkowski space.

At the end, the information contained in these features is combined using a feed-forward neural network that is specific to the given physics task. The entire model is trained jointly in a supervised training procedure.

## 3. Experiment and results

As an example application, we consider the identification of top-quark pair associated Higgs boson production, with the Higgs boson decaying into two bottom quarks ( $t\bar{t}H(H \rightarrow b\bar{b})$ ). The most difficult to distinguish background is top-quark pair production with an additional  $b\bar{b}$  pair ( $t\bar{t} + b\bar{b}$ ), for example due to gluon radiation.

In both cases, we have two top quarks, each decaying into a W boson and a bottom quark. We specifically consider the semi-leptonic decay channel, where the W boson originating from one



**Figure 1.** Schematic overview of our network architecture. The input four-vectors are first combined to particles and rest frames using trainable weights. The composite particles are then boosted into their corresponding rest frames and characteristic features are extracted from the boosted particles. These features are then used in a second stage for the specific physics analysis.

of the top quarks decays into two quarks, and the other one decays into a lepton and neutrino. This means we expect 6 jets, a lepton, and a neutrino in the final state. The neutrino can not be detected directly, but is reconstructed from the missing transverse energy by utilizing a  $W$  boson mass constraint.

The dataset used is generated with Pythia 8.2.26[3], and a parametric simulation of the CMS detector is performed using Delphes[4]. The reconstructed jets are matched to the partons of the generator and an unambiguous matching is required. In the following, the bottom quark jets from the hadronic and leptonic top decay branch are called  $b_{had}$  and  $b_{lep}$ , respectively. The two jets from the hadronically decaying  $W$  boson are denoted as  $q_1$  and  $q_2$ . Finally, the two remaining jets are labeled  $b_1$  and  $b_2$ .

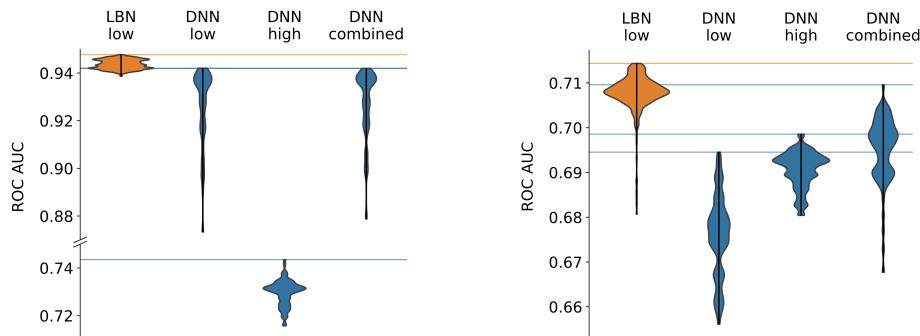
Our model is trained to distinguish signal from background events. To judge its performance, we also train a domain-unspecific neural network on the same separation task, which we will denote as DNN in the following. We provide this network with either the same final state four-vectors, with a set of high level variables dedicated for a  $t\bar{t}H(H \rightarrow b\bar{b})$  analysis, or with a combination of both.

We consider two different cases for our performance measurements. On the one hand, we use generator information to sort the final state four-vectors according to their matched partons. This allows us to isolate the study of the performance of our model from the task of identifying the input particles. On the other hand, sorting the input four-vectors by their transverse momenta corresponds to a more realistic physics use case. Detailed information on the network hyperparameters and set of high level variables for the DNN can be found in [5].

The results are summarized in fig. 2. The performance is measured by the integral of the receiver operating characteristic curve (ROC AUC). For the generator ordered data, we see on the one hand that the DNN already obtains the best results when using only low level variables, and does not improve by including additional high level information. On the other hand, the high level variables alone do not contain enough information to reach similar performance. This is because the high level variables are constructed to be independent of the input order, which is not

available in an analysis performed on data, and thus can not exploit this information. However, the LBN outperforms the domain-unspecific network and also shows more stable results.

When ordering the input four-vectors by their transverse momenta, we can see a very typical performance hierarchy for the DNN. The network can already extract useful information from the low level features. However, using the high level variables improves performance, and the combination of both yields the best result. Again, we obtain the best performance with the LBN, even though it has been trained only with the low level information.



**Figure 2.** Performance of the LBN and a feed-forward DNN, as measured by the ROC AUC. The LBN uses only four-vector information, while the DNN is supplied with four-vectors (low), dedicated high level variables (high), or both (combined). The input four-vectors are ordered either using generator information (left), or by their transverse momentum (right).

#### 4. Interpretation of network predictions

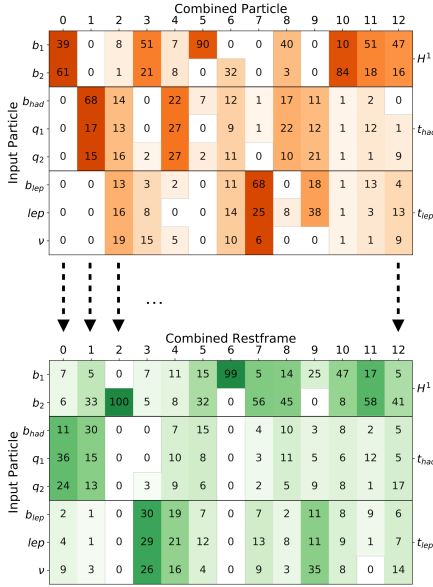
As multiple parts of our model have a direct physics interpretation, we can study them to gain insights into the internals of the network. All studies in the following will use generator sorted data in order to have a direct correspondence of input particles to physics objects.

Firstly, we can look at the weight matrix that is used to build the composite particles and rest frames. This matrix is shown in fig. 3. The model tends to group together particles originating from the same decay and thus approximately reconstructs the intermediate particles of the event. For example, particle 1 approximates the Higgs boson (or the  $b\bar{b}$  system in the case of  $t\bar{t} + b\bar{b}$ ).

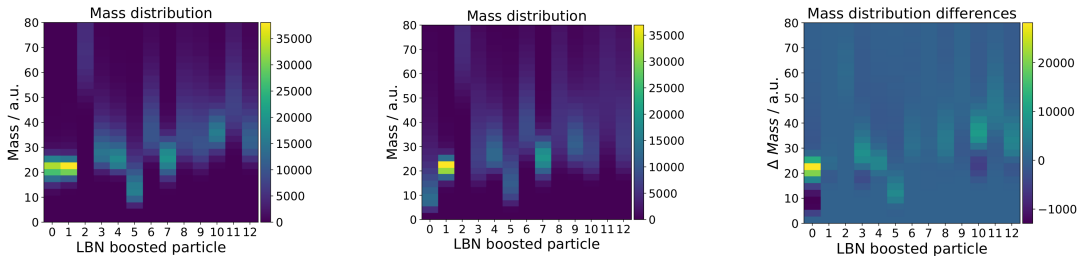
Another apparent feature is that there are strong contributions of the two particles originating from the Higgs boson. This matches our expectations because the main differences between the two processes should be related to these particles.

In addition, we can investigate the features that are constructed from the boosted particles. Figure 4 shows the masses of all 13 boosted particles and the difference in their distributions for signal and background events. The mass of particle 1 shows the strongest separation power. As can be seen in fig. 3, this particle approximates the Higgs boson.

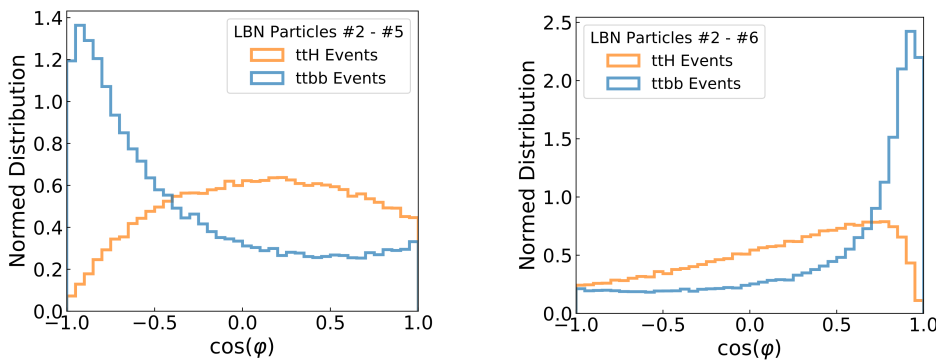
It should be noted that since the particle mass is a Lorentz invariant quantity, studying the particle masses only gives us insight into the combined particles, but not into their rest frames. For this, we can look at more complex features, such as the pairwise angles between particles. Two example features that exhibit strong separation power are shown in fig. 5.



**Figure 3.** Particle and rest frame combinations built by the LBN when trained on generator ordered data. The model focuses on the partons originating from the Higgs boson and approximately reconstructs the intermediate particles of the decay.



**Figure 4.** Masses of the boosted particles built by the LBN, shown for  $t\bar{t}H(H \rightarrow b\bar{b})$  events (left),  $t\bar{t} + b\bar{b}$  events (middle), and their difference (right). The largest difference occurs for particle 1, which is an approximate reconstruction of the Higgs boson, as can be seen in fig. 3.



**Figure 5.** Two examples of angular variables between two boosted particles in the LBN that show good separation power. The particle compositions are shown in fig. 3.

**5. Conclusion**

In this work, we introduced a two staged neural network architecture designed to autonomously characterize high energy physics collision events using only four-vector information. In the first stage, the Lorentz Boost Network, the four-vectors of the final state particles are combined

to create composite particles and rest frames. These particles are then boosted into their corresponding rest frames and characteristic features are extracted. The second stage uses these characteristic features for a specific physics analysis by combining them in a feedforward neural network.

We presented an application of our model to the task of identifying top-quark pair associated Higgs boson production events, and showed improved performance compared to using domain unspecific neural networks, even if these networks are provided with additional high level information.

Because multiple parts of our model have a direct physics interpretation, we can study what the network is learning by inspecting the weight matrices used to construct the combined particles and rest frames, as well as the constructed features. We find that the model forms physically meaningful particle combinations, which are then utilized to create well separating variables.

A further application of our model to the identification of boosted hadronically decaying top quarks has not been presented here, but is described as part of a method comparison in [6]. The source code of our implementation is publicly available [7].

### Acknowledgments

We wish to thank Jean-Roch Vlimant, Benjamin Fischer, Dennis Noll, and David Schmidt for valuable discussions. This work is supported by the Ministry of Innovation, Science and Research of the State of North Rhine-Westphalia and by the Federal Ministry of Education and Research (BMBF).

### References

- [1] Baldi P, Sadowski P and Whiteson D 2015 *Phys. Rev. Lett.* **114** 111801 (*Preprint* 1410.3469)
- [2] Sirunyan A M *et al.* (CMS) 2019 *JHEP* **03** 026 (*Preprint* 1804.03682)
- [3] Sjstrand T, Ask S, Christiansen J R, Corke R, Desai N, Ilten P, Mrenna S, Prestel S, Rasmussen C O and Skands P Z 2015 *Comput. Phys. Commun.* **191** 159–177 (*Preprint* 1410.3012)
- [4] de Favereau J, Delaere C, Demin P, Giammanco A, Lematre V, Mertens A and Selvaggi M (DELPHES 3) 2014 *JHEP* **02** 057 (*Preprint* 1307.6346)
- [5] Erdmann M, Geiser E, Rath Y and Rieger M 2018 Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering (*Preprint* 1812.09722)
- [6] Kasieczka G *et al.* 2019 The Machine Learning Landscape of Top Taggers (*Preprint* 1902.09914)
- [7] Erdmann M, Geiser E, Rath Y and Rieger M Lorentz Boost Network implementation [software], version 1.0.0 available from <https://git.rwth-aachen.de/3pia/lbn>