Geometrical Deep Learning for τ Identification

Kirill Bukin

Moscow Institute of Physics and Technology (MIPT)

Supervised by: Leonid Didukh, Dirk Kruecker, Isabell Melzer-Pellmann

Deutsches Elektronen-Synchrotron (DESY)

September 4, 2019

Abstract

The τ lepton reconstruction and identification is important for many searches. In this work, τ Isolation Discriminator capable of discrimination τ leptons from jets using geometrical deep learning techniques is presented. The network provides performance comparable to the performance of convolutional neural network-based DeepPF classifier, while being able to classify events 64 times faster.

Contents

1 Introduction

There is now a lot of interest to searches for different phenomena that include τ leptons in its signature. One of the examples of such searches is the decay of Higgs boson (H) into a pair of τ leptons which is the most sensitive channel for probing Higgs boson coupling to fermions. The observation of standard model (SM) Higgs boson decaying into a pair of τ leptons was reported in Ref. [1, 2]. Moreover, searches with τ leptons in the final state have high sensitivity to the production of additional Higgs bosons expected in the minimal supersymmetric standard model (MSSM). Examples of such searches can be found in Ref. [3, 4]. In addition, searches for different particles beyond the SM benefit significantly from any improvements made in τ lepton reconstruction and identification.

Therefore, different τ discriminators has been developed and are currently used. In our work, we present discriminator for tau identification based on geometrical deep learning and compare it to the official CMS τ discriminators.

1.1 The CMS Experiment

The Compact Muon Solenoid (CMS) is a general purpose detector operating at the Large Hadron Collider (LHC) at CERN. The overall layout and a slice of a detector is shown in the Fig. 1, 2. The CMS apparatus has an overall length of 22 m, a diameter of 15 m, and weighs 14 000 tonnes. The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid.

The particle-flow algorithm [5] aims to reconstruct and identify each individual particle in an event, with an optimized combination of information from the various elements of the CMS detector. The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [6].

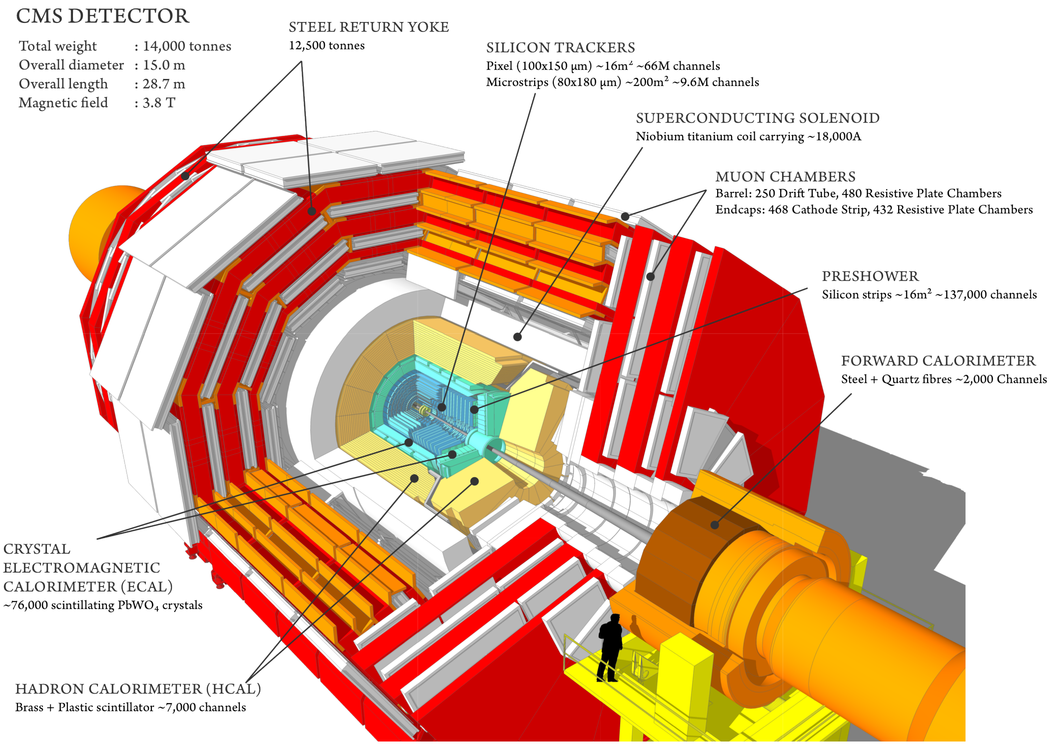


Figure 1: The CMS detector at the LHC

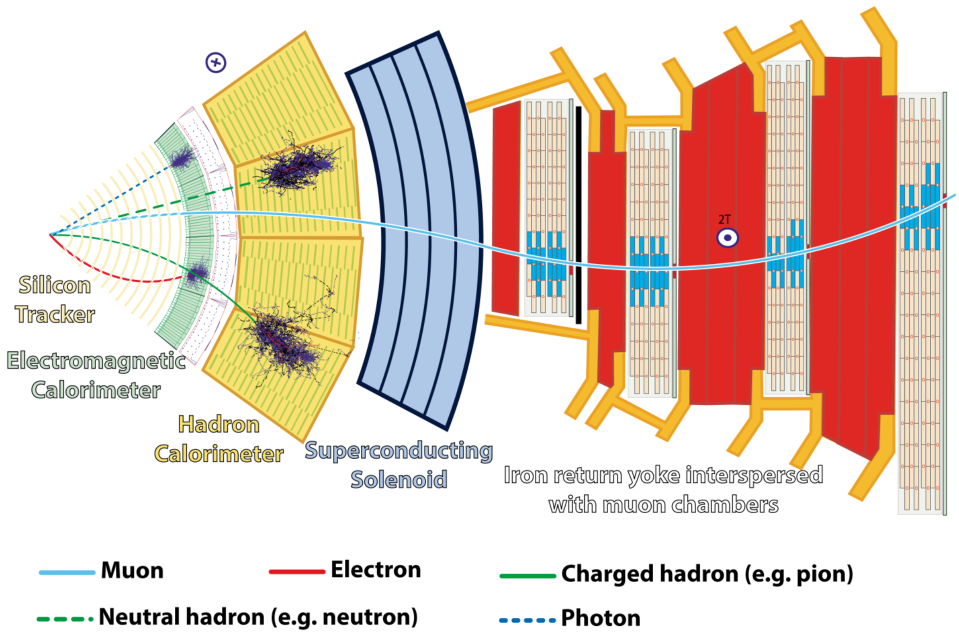


Figure 2: Slice showing CMS sub-detectors and how particles interact with them

1.2 τ lepton

The τ lepton is a third generation charged lepton, with a lifetime of $2.9 \cdot 10^{-13}$ s and a mass of 1.776 GeV [7]. It is the only lepton which is heavy enough to decay hadronically. Branching ratios of different τ decay modes are shown in Table 1. The branching ratio of hadronic τ lepton decays is 64.8 %, thus, improvement in efficiency of tau reconstruction in this mode can significantly improve overall tau reconstruction efficiency. Algorithm which is currently used for the reconstruction of hadronically decaying τ is based on particle-flow algorithm.

The hadronically decaying τ, denoted τ_h, are reconstructed using hadrons-plus-strips (HPS) algorithm [8, 9]. From neutral hadrons, charged hadrons, electron and photon PF candidates, HPS algorithm checks whether the combined topology is compatible with one of the τ_h decay modes. The main challenge in τ_h identification is then to distinguish it from quark and gluon jets arising in a large amounts in proton-proton collisions.

Table 1: Branching ratios (B) of different τ decay modes

Decay mode	Resonance	B(%)	
Leptonic decays		35.2	
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$			17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$			17.4
Hadronic decays		64.8	
$\tau^- \to h^- \nu_\tau$			11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$		25.9
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$		9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$		9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$			4.8
Other			3.3

2 Related Work

A number of different τ isolation discriminators were developed earlier and are currently used in physical analyses at CMS. Isolation sum and MVA discriminators were proposed in Ref. [9]. Cut-based isolation sum discriminator is computed by scale p_T sums of the charged particles and photons reconstructed using the PF algorithm with a cone centered around the τ_h:

$$\Delta\beta = 0.20 \times \sum p_T^{charged}(d_z < 0.2cm) \quad (1)$$

The isolation is then defined as the following:

$$I_{\tau_h} = \sum p_T^{charged}(d_z < 0.2cm) + \max(0, \sum p_T^{\gamma} - \Delta\beta) \quad (2)$$

MVA discriminator is based on boosted decision trees (BDT) and is combining the isolation and other differential variables sensitive to the τ lifetime. In recent years, two different discriminators based on deep neural networks (DNN) were developed. Colegrowe proposed Deep Particle Flow (DPF) network [10], which uses representation of PF candidates around reconstructed τ as 2D table, which is then fed into a convolutional neural network (CNN) consisting of 10 convolutional layers and 4 fully-connected layers. CMS Tau ID group developed DeepTau ID network, which is the current state-of-the-art τ discriminator. We compare the performance of our network to the MVA and DPF classifiers.

3 Network Architecture

Recent success of neural networks is partially attributed to the rapidly developing computational resources, the availability of big training data and to development of architectures that can effectively extract latent representations from data with Euclidean structure. Taking image data as an example, one can represent is as a regular grid in a Euclidean space. Convolutional Neural Networks are able to exploit shift-invariance, local connectivity, and compositionality of that data which allows it to extract meaningful local features and results in high performance of CNNs in image processing tasks: in most computer vision problems they by far outperform all the previously developed methods.

However, there are a lot of tasks where data has non-Euclidean structure. For example, data can be represented as a graph which can be irregular and have varying size, which means that most operations widely used for data on Euclidean domain, such as convolutions, can not be used for this data and one need specific operations to efficiently extract features from such a data. Recently there is a growing interest in extending deep learning approaches for such a data.

Another example of non-Euclidean data is a point cloud - scattered collections of points which comprise the output of the most 3D sensors. Qu and Gouskos, inspired by the notion of point cloud, proposed to represent jets as an unordered set of its constituent particles, effectively a "particle cloud" [11]. Such particle cloud representation of jets is efficient in incorporating raw information of jets and also explicitly respects the permutation symmetry. This enables them to achieve state-of-the-art performance on two representative jet tagging benchmarks. We adopt their approach for τ discrimination task.

3.1 Edge Convolution

Different ways to generalize convolutions for using on point clouds were developed. In our work we use the EdgeConv layer, which has been proposed by Y. Wang et. al in Ref. [12].

For EdgeConv operation, the point cloud is first represented as a directed graph - points are the vertices of the graph, and each vertex is connected with its k nearest neighbours. For each edge of the graph, the edge features is computed as a $e_{ij} = h_\Theta(x_i, x_j)$, where x_i and x_j are the points connected by the edge, and h_Θ is a function parametrized by parameter vector Θ. Function h_Θ is usually implemented as a multi-layer perceptron (MLP). EdgeConv operation is defined by applying a symmetric aggregation operation $\square$ (e.g. $\sum$ or max)(see Fig. 3):

$$x_i' = \square h_\Theta(x_i, x_j) \tag{3}$$

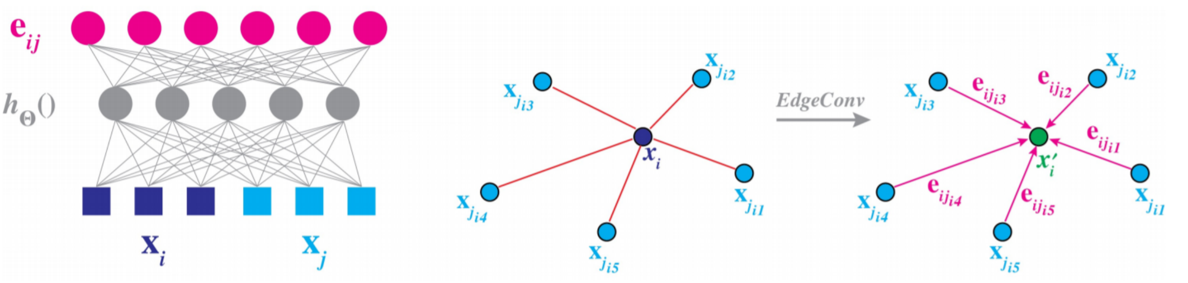


Figure 3: Left: An example of computing an edge feature. Right: Visualisation of Edge-Conv operation.

There are different ways in which arguments can be fed into MLP, and in our work we use asymmetric edge function, which explicitly combines global shape structure with local neighbourhood information:

$$h_\Theta(x_i, x_j) = \bar{h}_\Theta(x_i, x_j - x_i) \tag{4}$$

4 Evaluation

4.1 Monte Carlo Samples

For training and testing of the neural network, the samples from 2016 MC was used. True τ were selected from Drell-Yan MC samples, fake ones from W+jet MC samples. In both cases, only decays into one charged hadron (decay mode 0), one charged hadron and a neutral pion (decay mode 1), or three charged hadrons (decay mode 10) are selected (see Fig. 4). Reweighting is applied to samples in order to flatten in tau p_T spectrum.

4.2 Results

Final architectures used in performance evaluation are shown at Fig. 5. Both networks consist of few EdgeConv layers followed by channel-wise global mean pooling and fully connected classifier layers. Number of nearest neighbours used (denoted as k) and output sizes for all the MLP used in models (both in EdgeConv layers and in classifier) is also shown at Fig. 5. For ECN with 3 layers, dynamic graph is used - knn-graph is recomputed before each EdgeConv layer. Networks were implemented using PyTorch

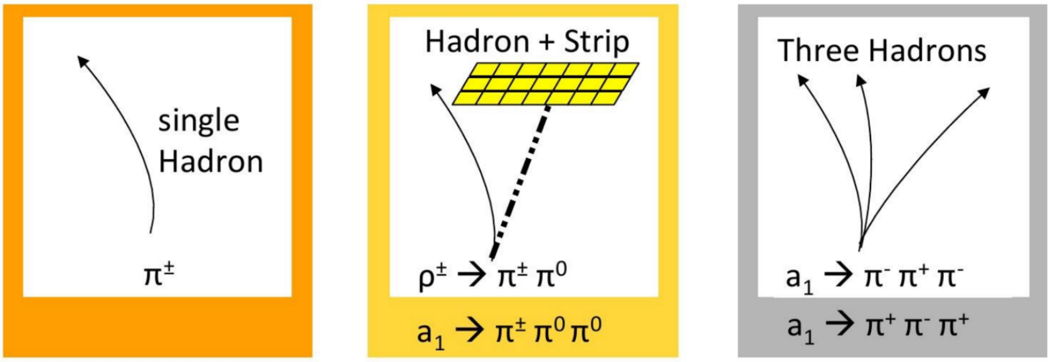


Figure 4: Tau decay modes used in our work

Table 2: Performance comparison for τ / jet discrimination

Architecture	ROC AUC score
MVA	0.792
DPF	0.863
DPF + MVA	0.892
ECN (1 layer)	0.883
ECN (3 layers)	**0.898**

Geometric library [13]. Neural networks were trained with stochastic gradient descent (SGD) optimiser using cosine annealing learning rate [14] with initial learning rate of 0.1 for 50 epoch with batch size equal 512. The cosine annealing LR is defined as follows:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + \cos(\frac{T_{cur}}{T_i}\pi)), \tag{5}$$

where T_{cur} is the number of current epoch and T_i is the specified number of epochs.

The ROC curves for MVA, ensemble of MVA and DPF and ECN are shown at the figure 6. The results are summarised in the Table 2.

Table 3: Number of parameters and inference time per object for different models

Architecture	Number of parameters	Time [ms]
DPF	8838945	166
ECN (1 layer)	**49283**	**0.6**
ECN (3 layers)	223939	2.6

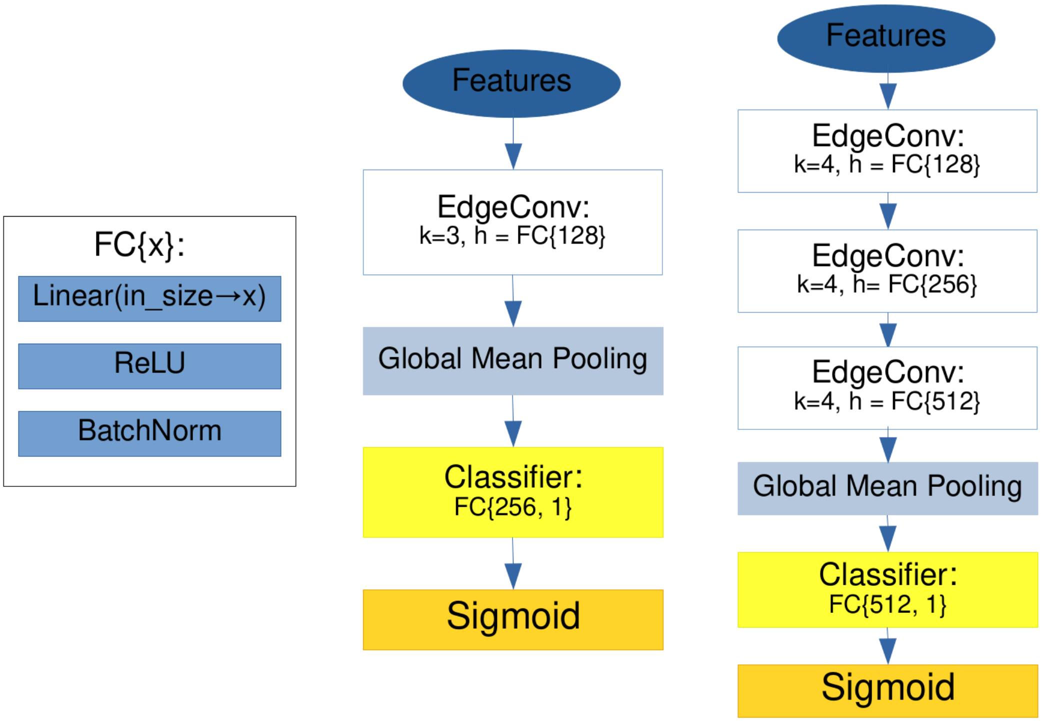
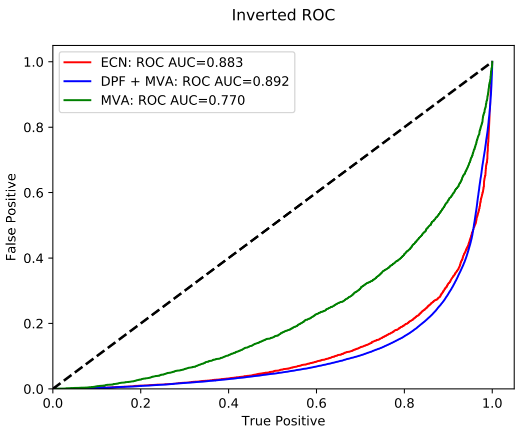
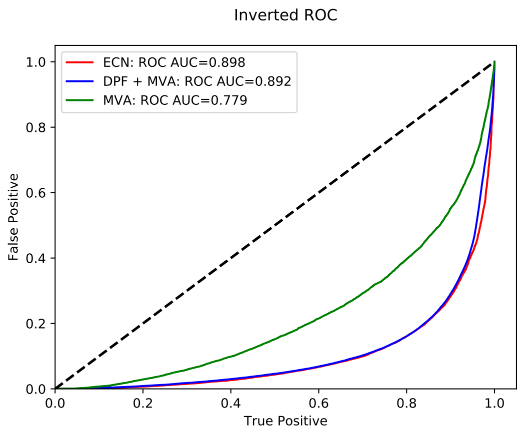


Figure 5: **Left:** Single fully-connected layer; **Center:** ECN with 1 conv layer; **Right:** ECN with 3 conv layers



(a) ECN with 1 EdgeConv layer



(b) ECN with 3 EdgeConv layers

Figure 6: Inverted ROC curves

5 Model complexity

Another important characteristic for network usage in analysis is the time needed to process one event. As most of the CMS analyses use a lot of data, tau tagging with highly time-consuming classifier would be unreasonable. One of the simplest ways to reduce time consumption is to reduce number of learnable parameters in the network thus reducing number of operations required to produce the output.

Comparement of the number of learnable parameters and the amount of time for evaluation per event for DPF network and our network can be found in the Table 3. Time was measured on CPU Intel Xeon E5-2650 using one thread and batch size of 2048.

6 Conclusion

In this work particle cloud approach was applied to the tau discrimination task. Geometrical neural network with particle cloud data representation achieved performance close to the one of conventional deep convolutional neural network while having much smaller number of parameters and working faster. Further improvement of performance can be achieved by optimising hyperparameters of the model, e.g. number of layers, batch size, learning rate and the number of nearest neighbours used to make knn-graph. In addition, the same architecture can be used to make multiclass classifier to discriminate τ_h from jets, e and μ.

Acknowledgments

Author would first like to express his gratitude to his supervisors Leonid Didukh, Dirk Kruecker and Isabell Melzer-Pellmann for the interesting project to work on and for the assistance received while working on the project. Then, author would like to thank Summer Student Programme organizing team and especially Olaf Benhke for providing an opportunity to participate in the program and gain enjoyable and valuable experience.

References

[1] ATLAS and CMS Collaborations. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. 2016.

[2] CMS Collaboration. Observation of the Higgs boson decay to a pair of tau leptons with the CMS detector. 2017.

[3] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions. 2014.

[4] CMS Collaboration. Search for a charged Higgs boson in pp collisions at sqrt(s) = 8 TeV. 2015.

[5] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12:P10003, 2017.

[6] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[7] Particle Data Group. Review of Particle Physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[8] CMS Collaboration. Performance of tau-lepton reconstruction and identification in CMS. 2011.

[9] CMS Collaboration. Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS. 2015.

[10] Search for direct τ slepton pair production in proton-proton collisions at $\sqrt{s} =$ 13 TeV. Technical Report CMS-PAS-SUS-18-006, CERN, Geneva, 2019.

[11] Huilin Qu and Loukas Gouskos. ParticleNet: Jet Tagging via Particle Clouds, 2019.

[12] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic Graph CNN for Learning on Point Clouds, 2018.

[13] Matthias Fey and Jan E. Lenssen. Fast Graph Representation Learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[14] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts, 2016.