Analysis of Different Loss Function for Designing Custom CNN for Traffic Sign Recognition

Virendra Patel [a], Sanyam Shukla [a] and Manasi Gyanchandani [a]

[a] *Maulana Azad National Institute of Technology, Bhopal, India*

Abstract
Traffic sign recognition has got important due to emerging AI based vehicles, one of the tasks required by these vehicles is automatic traffic sign recognition. CNN is a popular at tool which is deployed for image classification tasks. Several CNN based solutions have been proposed for traffic sign recognition, this work also develop a custom CNN for traffic sign recognition. The choice of loss function is a part of designing custom neural network. This work analyses different loss function for traffic sign recognition. The GTSRB dataset (German Traffic Sign Recognition Benchmark) is used in this study to conduct the trials. The outcomes of our experiments suggest that our prediction is right. For TSR, the suggested classifier beats previous techniques. This study attempts to address such an element by training the network to react in line with a rich database consisting 39,209 color photographs spanning 43 distinct kinds of signalized intersections, as well as system testing with 12,630 samples and reaching 99.81% accuracy.

Keywords
Loss function, Image recognition, GTSRB, TSR, Regularized CNN

1. Introduction

Traffic sign recognition is an area of interest for researches due to emerging advanced small vehicles. These vehicles have feature like voice command, Driver assistance system, Autopilot, fuel measurement and alert. The automatic feature like autopilot, driver assistance requires the vehicle to identify the traffic sign [1]. Traffic sign recognition has following based steps:

1. Traffic sign detection
2. Traffic sign recognition

Several approaches have been proving for traffic sign recognition problems.

- Classification of approaches based on neural network type:
- Shallow neural networks for traffic sign recognition:

Color is an important part of the data supplied to a user in needed to guarantee that the goals of the speed limit signs are realized. As a consequence, in important to stand out, road signs and its color are selected to clash with the natural setting or surroundings. The recognition of these indicators in exterior images taken from a passing car will aid the operator in deciding the best feasible in the shortest amount of time, leading to fewer fatalities, lower emissions, and improved security [2].

- Deep neural networks for traffic sign recognition:

WINS-2022: Workshop on Intelligent Systems, April 22 – 24, 2022, Chennai, India.
EMAIL: virendra23aitr@gmail.com (Virendra Patel)
ORCID: 0000-0001-8263-0468 (Virendra Patel)

The 'deep' in a 'Deep CNN' refers to the network's layers. In a standard CNN, 5–10 or even more feature learning layers are usual. Networks of more than 50–100 layers are common in modern topologies utilized in cutting-edge applications [3].

1. Classification of traffic sign recognition using feature extraction method:

- **Wavelet transform**: A wave sequence is a set of basic functions created by a wavefront that represents the value of a rectangular shape. A ripple is a pulse fluctuation that pulsates from zero-to-zero current and from zero to low resistance at a frequency that pulsates from zero to low resistance [6].

- **Fast Fourier Transform (FFT)**: A signal is then converted from the duration region to the spectral domain and return using the Fourier Transform. It converts capacity-dependent units into values that are based on geographic frequency. If the organised pairs defining the input data value are scattered evenly in their response variable, Fast Fourier Transform is considered Discrete Fourier Change [7].

- **Curvelet transform**: Wavelet coefficients generalise the Transformation function by employing a foundation that incorporates both location and geometric characteristics. The use of kernels that are also localised in position when directing harmonic transformations for three-dimensional inputs goes even further [8].

- **Gabor filter**: The Gabor filter is a texture analysis linear filter that searches an image for certain frequency components in specific directions within a constrained region around the research point or region [9].

- **CNN as a feature extraction**: CNN is a multilayer perceptron that collects input feature representations and classifies them to use another neural net. The extracting features network uses the input image. The neural network uses the feature extraction signals to classify the data [10].

This work develops a custom CNN based solution for traffic sign recognition, one of the important factors for designing custom CNN is choice of loss function. This work studies following loss function for the design of custom CNN:

1. KL-Divergence loss function
2. Poisson loss function
3. Categorical cross entropy loss function
4. Categorical hinge loss function
5. Binary cross entropy loss function

2. Related Work

1. **TsingNet CNN Model:** The TsingNet Convolution Layer can be used to train magnitude and situational features in order to recognize and categories undersized and limited highway signage in real time. TSingNet creates a worldwide active learning interest bidirectional multi-layer perception neural structure that combines underneath and leading network nodesto flow low-, mid-, and higher ambient interpretations [10].

2. **DeepThin CNN Model:** Because each convolution operation in this technique involves around 43 variables, we may create our CNN architecture without requiring a GPU.Using the massive GTSRB sample, we begin assessing the suggested architect's productivity. The suggested architecture outperformed current traffic sign approaches, which had at least 5 times less characteristics in each end-to-end learning link [11].

3. **ResNet50:** In a ResNet variation, ResNet50 has 48 Convolution operations, 1 MaxPool layer,

and 1 Average Pooling layer. There are 3.8 x 109 bobbing processes in it. We spent a lot of time looking at the ResNet50 architecture because it's a prevalent ResNet architecture. ResNet-50 is, in fact, a 50-layer neural network. We can import a from Imagenet, which has been learned from over a thousand photos and was before iteration of the structure. The algorithm, which previously recognised 1000 different item kinds in pictures, has been improved [12].

4. **VGG-16:** The number 16 in the VGG-16 designation refers to the assumption that there are a total of 16 layers, each with a different weight. With almost 138 million parameters, this is a massive network. The major issue was that it was a rather large network in terms of the number of variables to be learned [13].

3. Proposed Work

The GTSRB dataset is a freely available sample that is being utilised in a research investigation. The dataset is pre-processed before being split into training and testing sets in this study. Batch Normalization and Dropout regularised techniques, as well as the Adam optimizer, were used. To analyse the progress of the training and validation sessions, a graph for accuracy and loss function is generated. The GTSRB training set consists of 39,209 image characteristics separated into 43 classes, whereas the testing set consists of 12,630 samples.

Five different loss functions KL-Divergence loss function, Poisson loss function, Categorical cross entropy loss function, Categorical hinge loss function and Binary cross entropy loss function uses for analysis of custom CNN model.

We employed four convolution layers, two MaxPooling layers, one Flatten layer, one Dense layer, twoDropout layers, and three Batch Normalization Regularization techniques in our customised CNN model. Convolutional layers are used to add filters to the original image or even a different set of inputs within a deep network. The majority of subscriber variables are stored within the system.

The most important qualities are the frequency and size of kernels. MaxPooling is a method of filteringthat selects the largest component from the feature space covered by the filtering. Like a result, the max-pooling layer's output is an excellent approach to identify the dataset's more notable traits.

The technique of consolidating data in one place collection for it at a high level is known as flattening. We compress the convolution theory's output to generate a unique long influence the structure. It's also linked to the final categorization approach, also known as the completely layering framework. In a CNNdesign, a dense layer is one that is tightly related to the one before it, meaning that the layer's cells are linked to every cell in the layer before it. This is by far the most commonly used layer in convolutionalneural network systems.

The relu and softmax activation functions were used in our customised CNN model. The activation function of a deep neural network can be defined and added to aid in the acquisition of complex data patterns. The activation function, in contrast to the nerve cell image we see in our heads, is in a better position to decide what should be given to another brain at the end of the process. The usage of ReLU prevents the amount of computing power required to run the neural network from rising exponentially.

As the size of the CNN grows, the computational complexity of integrating more ReLU velocity grows. The ReLU has been the most extensively used bias vector in the planet right now. It has been used in practically all deeper neural network models analytical approaches ever since. To determine a regression analysis conditional probability, the softmax function is employed as the perceptron in the hidden layers of normal neural analysis. Softmax is used as the activation function for interclassification challenges needing classifiers on more than two words. To avoid overfitting our CNN model, we used Adam optimizer.

Adam is a deep learning training method that uses a different algorithm than gradient descent to create deep learning models. By integrating the best aspects of the AdaGrad and RMSProp approaches, Adam develops an optimization strategy for noisy situations with dense gradients.

3.1. Architecture of Proposed CNN Model

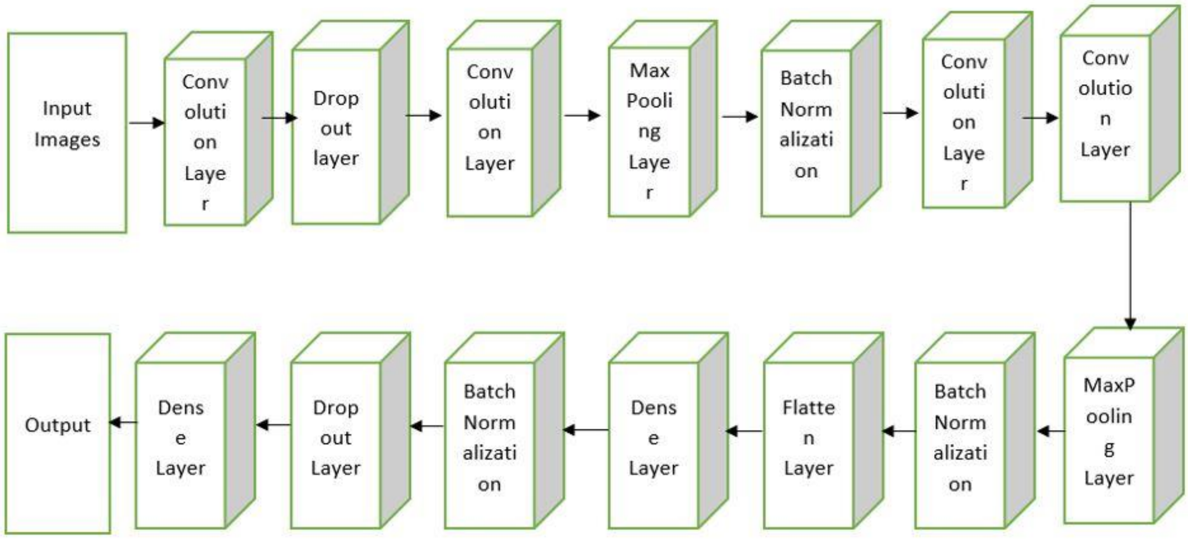

Figure 1: Customized CNN Model

Layer	Layer Type	Feature Map Number	Convolution Kernel Size	Feature Map Size
1	Convolution Layer	28	3 x 3	32 x 32
2	Dropout Layer	28	-	32 x 32
3	Convolution Layer	26	3 x 3	64 x 64
4	MaxPooling Layer	13	2 x 2	64 x 64
5	Batch Normalization	13	-	64 x 64
6	Convolution Layer	11	3 x 3	128 x 128
7	Convolution Layer	9	3 x 3	256 x 256
8	MaxPooling Layer	4	2 x 2	256 x 256
9	Batch Normalization	4		256 x 256
10	Flatten	-	-	512 x 512
11	Dense	-	-	512 x 512
12	Batch Normalization	-	-	512 x 512
13	Dropout Layer	-	-	512 x 512
14	Dense Layer	43	-	-

Figure 2: Different layers in Customized CNN Model

3.1.1. Choice of loss function for above CNN model

This work determines the optional choice of loss function for the above CNN by experimentation the loss function consider for the design of custom CNN’s are:

1. **KL-Divergence loss function**: Kullback-Leibler divergence is a score that determines how far one probability distribution differs from another. Jensen-Shannon divergence is an extension of KL divergence that determines a symmetric scoring and proximity measurement between two probabilistic. The KL divergence is the minus summation of every show's probability in P multiplied by the log of the show's probability in Q over the probability of the event in P [14].

$$\boldsymbol{KL\ (P\ ||Q) = \ -sum\ x\ in\ X\ (x) * \log\left(\frac{Q(x)}{P(x)}\right)} \qquad (1)$$

2. **Poisson loss function**: When modelling count data, the poisson loss function is utilised for regression. The poisson distribution is used for data. For example, next week's consumer churn. The loss manifests itself as:

$$\boldsymbol{L(y,\hat{y}) = \frac{1}{N}\sum_{i=0}^{N}(\hat{y}_i - y_i\, log\,\hat{y}_i)} \qquad (2)$$

Where ŷ is the predicted expected value.

Minimizing the Poisson loss is the same as maximising the data's likelihood under the assumption that the target is drawn from a Poisson distribution that is conditioned on the input [15].

3. **Categorical cross entropy loss function**: Categorical cross entropy is a loss function in multi-class text categorization. All of those are issues where an instance can only fit into one of several classes, and the system have to choose one. Its basic goal is to compare two probability distribution function and identify the difference [16].

$$\boldsymbol{CE = \ -log\left(\frac{e^{s_p}}{\sum_{j}^{c} e^{s_j}}\right)} \qquad (3)$$

4. **Categorical hinge loss function:** The hinges loss is an error term used it to training classifications in computer vision. For "maximum-margin" classification, the hinge loss is utilised, most notably for support vector machines (SVMs). The hinge loss of the prediction y is defined as for an expected output t = $\pm$1 and a classifier score y [17].

$$\boldsymbol{l(y) = max(0, 1 - t.y)} \qquad (4)$$

5. **Binary cross entropy loss function:** The current class result, which might be either 0 or 1, is checked to every one of the predicted chances. The rating would then be computed, with chances being penalised based on how far they deviate from of the projected value. This indicates whether near however far a number is to the true value [18].

$$\boldsymbol{logloss = \ -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\, log(p_{ij})} \qquad (5)$$

4. Dataset

This dataset was obtained from Kaggle Library by GTSRB (German Traffic Sign Recognition Benchmark), and it is freely available. A total of 51839 images are used, with 39209 and 12630 images accounting for around 70% and 30% of the training and testing sets, respectively. The photos are 512*512 pixels in size and 32*32 pixels in size, respectively.

5. Experimental Result and Analysis

Table 1: Comparison of result between different loss functions in proposed CNN mode on GTSRB Dataset

Loss function	Accuracy
KL-Divergence	99.36%
Poisson	99.41%
Categorical Cross entropy	99.56%
Categorical hinge	99.76%
Binary Cross entropy	99.81%

Table 2: Comparison of result between different CNN mode on GTSRB Dataset

CNN Model	Accuracy
SCN model [1]	97.81%
MLADA Model [5]	94.50%
BBAS model [3]	98.12%
CoDefend model [4]	95.26%
Proposed CNN model	99.81%

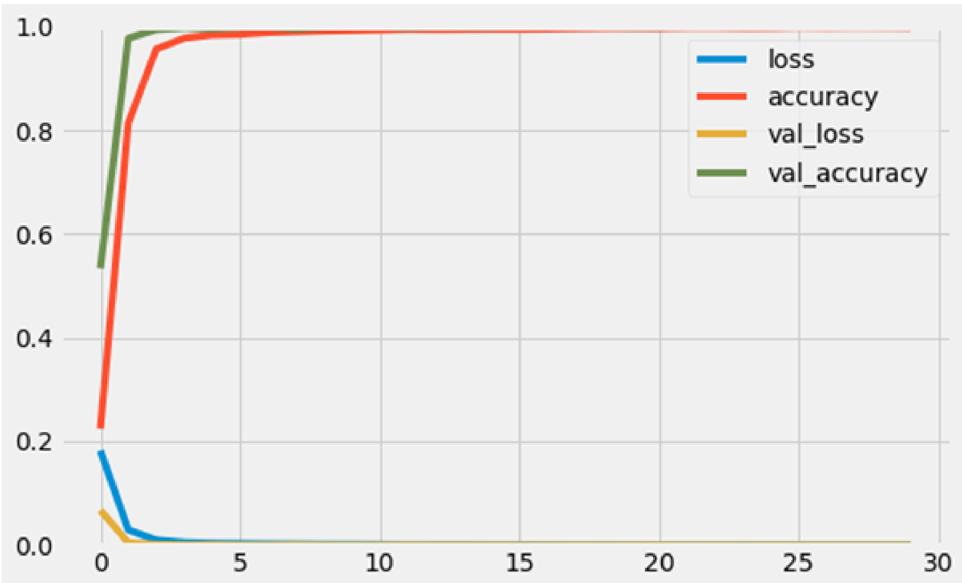

Figure 3: Training and validation accuracy of GTSRB dataset with binary cross entropy loss function

6. Conclusion

This paper introduces a new traffic sign identification system based on a customized neural network. This paper investigates the use of different loss functions such as KL-Divergence loss function, Poisson loss function**,** Categorical cross entropy loss function**,** Categorical hinge loss function and Binary cross entropy loss function to improve TSR performance. In noisy photos and when the weather is not clear, our proposed CNN model outperforms the existing approach. Future study will include a thorough examination of CNN design and learning algorithms in order to improve performance even more. Future study will also include the development of image annotation techniques to improve the amount of framing data available, resulting in a higher TSR.

7. References

[1] Shi, G., Zhang, Z., Gao, D., Lin, J., Xie, X. & Liu, D. Knowledge- guided semantic computing network. *Neurocomputing*. **426** pp. 70-84 (2021)

[2] Fleyeh, Hasan, and Al-Hasanat RM Bin Mumtaz. "Adaptive shadow and highlight invariant colour segmentation for traffic sign recognition based on Kohonen SOM." *Journal of Intelligent Systems* 20.1 (2011): 15-31.

[3] Shen, J. & Robertson, N. BBAS: Towards large scale effective ensemble adversarial attacks against deep neural network learning. *Information Sciences*. **569** pp. 469-478 (2021)

[4] Liu, Q., Chen, L., Jiang, H., Wu, J., Wang, T., Peng, T. & Wang, G. A collaborative deep learning microservice for backdoor defenses in Industrial IoT networks. *Ad Hoc Networks*. pp. 102727 (2021)

[5] Fang, Y., Xiao, Z. & Zhang, W. Multi-layer adversarial domain adapta- tion with feature joint distribution constraint. *Neurocomputing*. **463** pp. 298-308 (2021)

[6] Yadav, A., Sengar, N., Issac, A. & Dutta, M. Image processing-based acrylamide detection from fried potato chip images using continuous wavelet transform. *Computers And Electronics In Agriculture*. **145** pp. 349-362 (2018)

[7] Jimenez, P., Basc o´n, S., Moreno, H., Arroyo, S. & Ferreras, F. Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies. *Signal Processing*. **88**, 2943-2955 (2008)

[8] Devapal, D., Hashna, N., Aparna, V., Bhavyasree, C., Mathai, J. & Soman, K. Object detection from SAR images based on curvelet despeckling. *Materials Today: Proceedings*. **11** pp. 1102-1116 (2019)

[9] Park, J. & Kim, K. Design of a visual perception model with edge- adaptive Gabor filter and support vector machine for traffic sign detec- tion. *Expert Systems With Applications*. **40**, 3679-3687 (2013)

[10] Liu, Y., Peng, J., Xue, J., Chen, Y. & Fu, Z. TSingNet: Scale-aware and context-rich feature learning for traffic sign detection and recognition in the wild. *Neurocomputing*. **447** pp. 10-22 (2021)

[11] Haque, W., Arefin, S., Shihavuddin, A. & Hasan, M. DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements. *Expert Systems With Applications*. **168** pp. 114481 (2021)

[12] Deshpande, A., Estrela, V. & Patavardhan, P. The DCT-CNN-ResNet50 architecture to classify brain tumours with super-resolution, convolu- tional neural network, and the ResNet50. *Neuroscience Informatics*. pp. 100013 (2021)

[13] Lin, D., Li, Y., Prasad, S., Nwe, T., Dong, S. & Oo, Z. CAM-guided Multi-Path Decoding U-Net with Triplet Feature Regularization for Defect Detection and Segmentation. *Knowledge-Based Systems*. **228** pp. 107272 (2021)

[14] Wang, Shuai, Yanmin Qian, and Kai Yu. "Focal KL-divergence based dilated convolutional neural networks for co-channel speaker identification." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.

[15] Li, C. P., and H. B. Hao. "E-Bayesian estimation and hierarchical Bayesian estimation of Poisson distribution parameter under entropy loss function." *IJAM* 49 (2019): 369-374.

[16] Ho, Yaoshiang, and Samuel Wookey. "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling." *IEEE Access* 8 (2019): 4806-4813.

[17] Kavalerov, Ilya, Wojciech Czaja, and Rama Chellappa. "A multi-class hinge loss for conditional gans." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021.

[18] Sypherd, Tyler, et al. "A tunable loss function for binary classification." *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019.

[19] Campbell, A., Both, A. & Sun, Q. Detecting and mapping traffic signs from Google Street View images using deep learning and GIS. *Computers, Environment And Urban Systems*. **77** pp. 101350 (2019)

[20] Li, C., Qu, Z., Wang, S. & Liu, L. A method of cross-layer fusion multi-object detection and

recognition based on improved faster R-CNN model in complex traffic environment. *Pattern Recognition Letters*. **145** pp. 127-134 (2021)

[21] Özkaya, U., Öztürk, Ş., Melgani, F. & Seyfi, L. Residual CNN+ Bi- LSTM model to analyze GPR B scan images. *Automation In Construc- tion*. **123** pp. 103525 (2021)

[22] Mou, L., Zhou, C., Zhao, P., Nakisa, B., Rastgoo, M., Jain, R. & Gao,W. Driver stress detection via multimodal fusion using attention-based CNN-LSTM. *Expert Systems With Applications*. **173** pp. 114693 (2021)

[23] Guo, S., Guo, D., Chen, L. & Jiang, Q. A L1-regularized feature selection method for local dimension reduction on microarray data. *Computational Biology And Chemistry*. **67** pp. 92-101 (2017)

[24] Gorba, C. & Cebrian, J. The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal Of Phonetics*. **88** pp. 101094 (2021)

[25] Skourt, B., El Hassani, A. & Majda, A. Mixed-pooling-dropout for convolutional neural network regularization. *Journal Of King Saud University-Computer And Information Sciences*. (2021)