Road Traffic Sign Classification under Challenging Conditions

Weixi FENG, UG; Prof. Kenneth Lam, Supervisor

The Hong Kong Polytechnic University

Department of Electronic and Information Engineering

- Introduction
- Proposed network
- Dataset
- Data augmentation
- Experiment results
- Defects and Future works
- References

Introduction

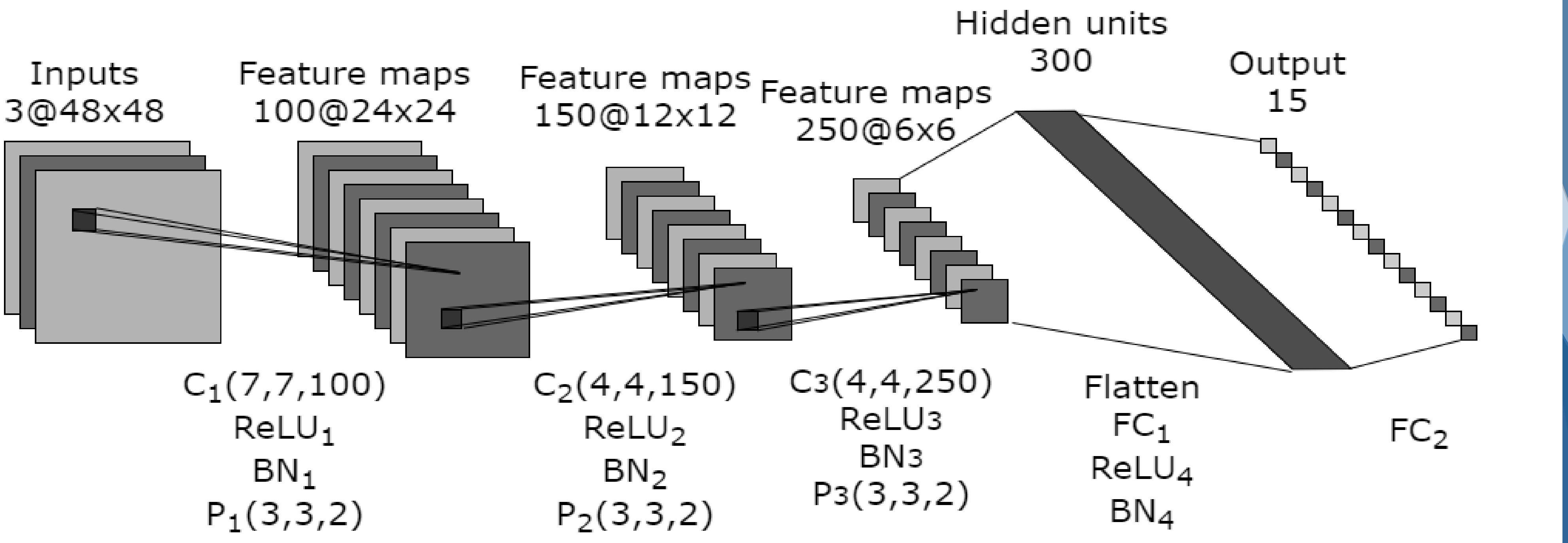
- State of the art recognition algorithms have nearly perfect performances on most of the existing traffic sign datasets.

- All these dataset either do not include challenging conditions or fail to emphasize the relationship between challenging environmental conditions and algorithmic performances.

Related work

- Before the widespread of Convolutional Neural Network(CNN), classification is completed based on conventional methods such as Support Vector Machine.
- However, none of the conventional methods outperform the best Deep Learning based method, which won the first place on GTSRB competition several years ago. Recent CNN-based methods include creating ensemble of CNNs, multi-scale CNN and much deeper networks.

Proposed network



$C(w, w, k)$: Convolutional layer with k filters of size $w \times w$
ReLU: ReLU activation layer
BN: Batch Normalization layer
$P(w, s)$: Max-pooling layer with kernel size w and stride s

Proposed network

- Batch normalization

Without batch normalization, we must be very careful in parameter initialization and learning rate selection to avoid internal covariate shift. Batch normalization could prevent such problem as well as overfitting[4].It allows us to set a large learning rate without the concern that saturating non-linearity would happen.

Input: values of x over minibatch: $\mathcal{B} = \{x_{1...m}\}$

Output:

$$\mu_{\mathcal{B}} = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{minibatch mean}$$

$$\sigma_{\mathcal{B}}^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{minibatch variance}$$

$$\widehat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{normalize}$$

$$y_i = \gamma \widehat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i) \qquad \text{scale and shift}$$

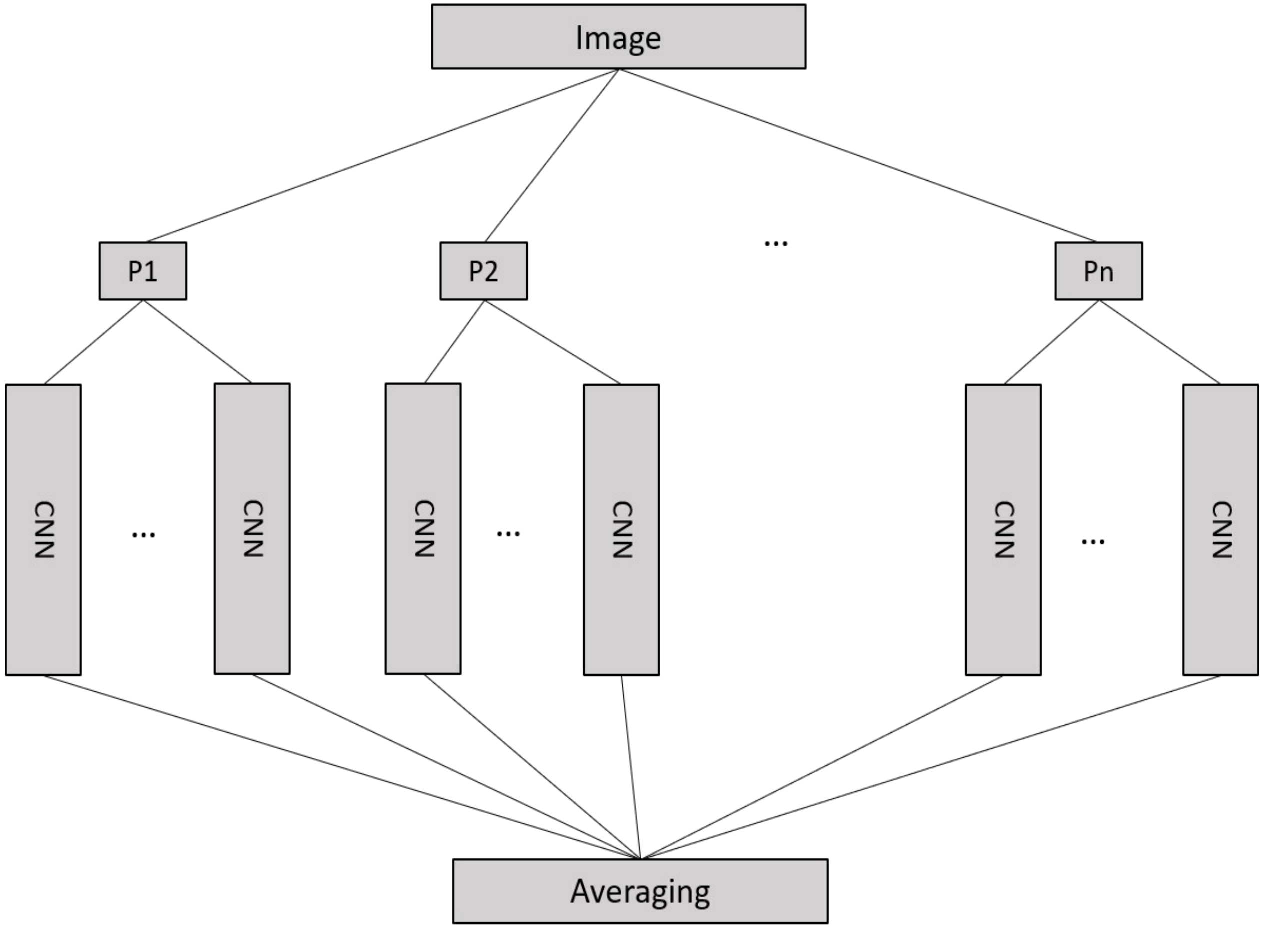
- Activation function

The advantages of using ReLU, compared to tanh, are that ReLU can increase the sparsity and solve the ‘vanishing gradient’ problem [5]

Ensemble of CNNs

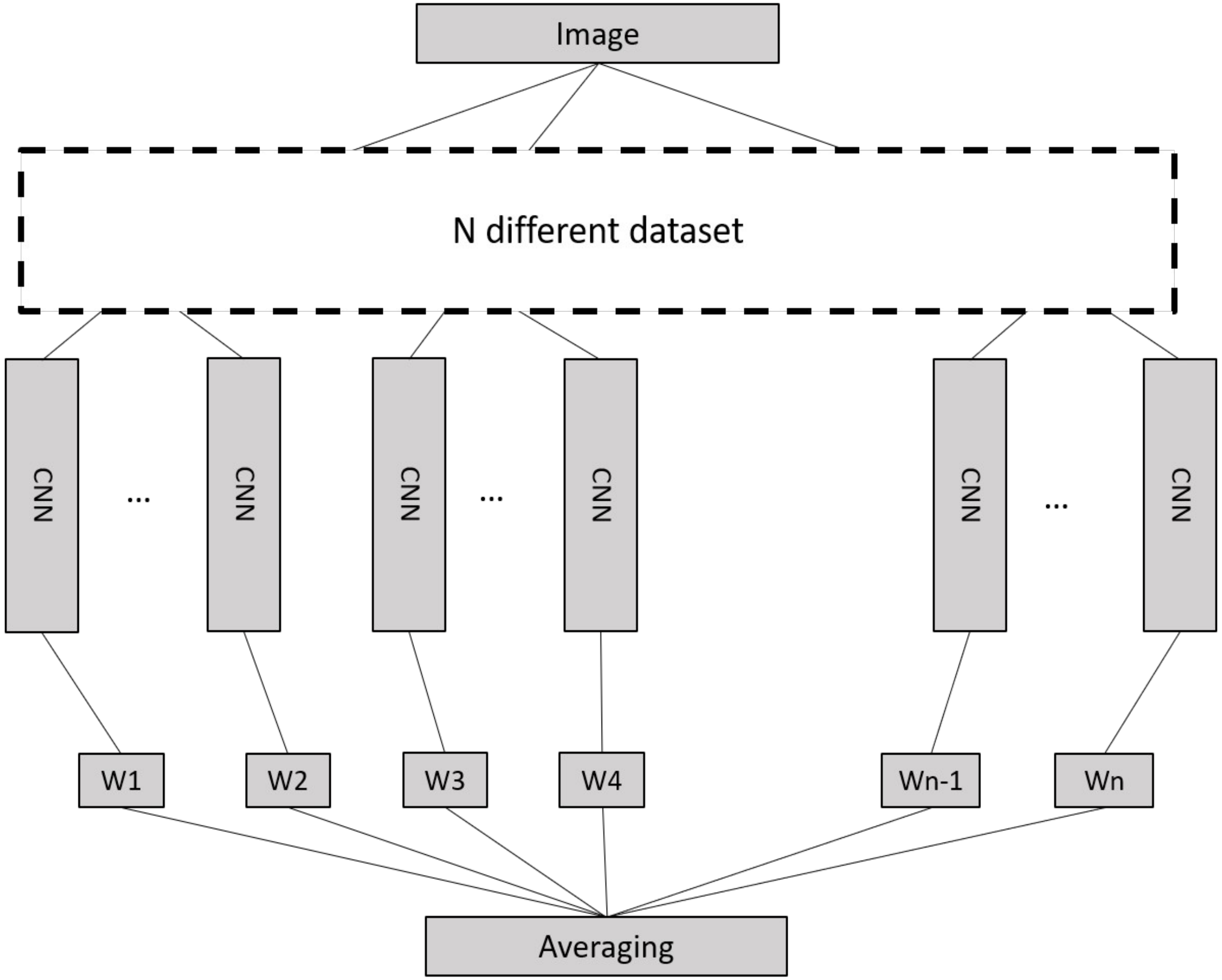
- **Theory**: If the errors of P different models have zero mean and are uncorrelated, the average error might be reduced by a factor of P simply by averaging the P models[7].
- **Problems to think:**
- In practice, errors of models trained on similar data tend to be highly correlated.
- Whether to optimize the combination of outputs of various models or not?
- Additional Data?
- Highly generalized to the unseen testing data?

Ensembles



[2] D. Cireşan, U. Meier, J. Masci and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification", *Neural Networks*, vol. 32, pp. 333-338, 2012.

Ensembles



$$\mathrm{E} = \sum_{k=1}^{K} \left\| y_i - \sum_{i=1}^{5} w_i \vec{S}_i \right\|$$

$$w_i \in [0.150, 0.250], \quad \sum_{i=1}^{5} w_i = 1$$

Dataset

- The VIP Cup Dataset has:
  - 49 real scenes + 49 unreal scenes
  - 14 types of traffic signs
  - 12 challenging conditions
  - 5 levels of severity

level 1 level 2 level 3 level 4 level 5

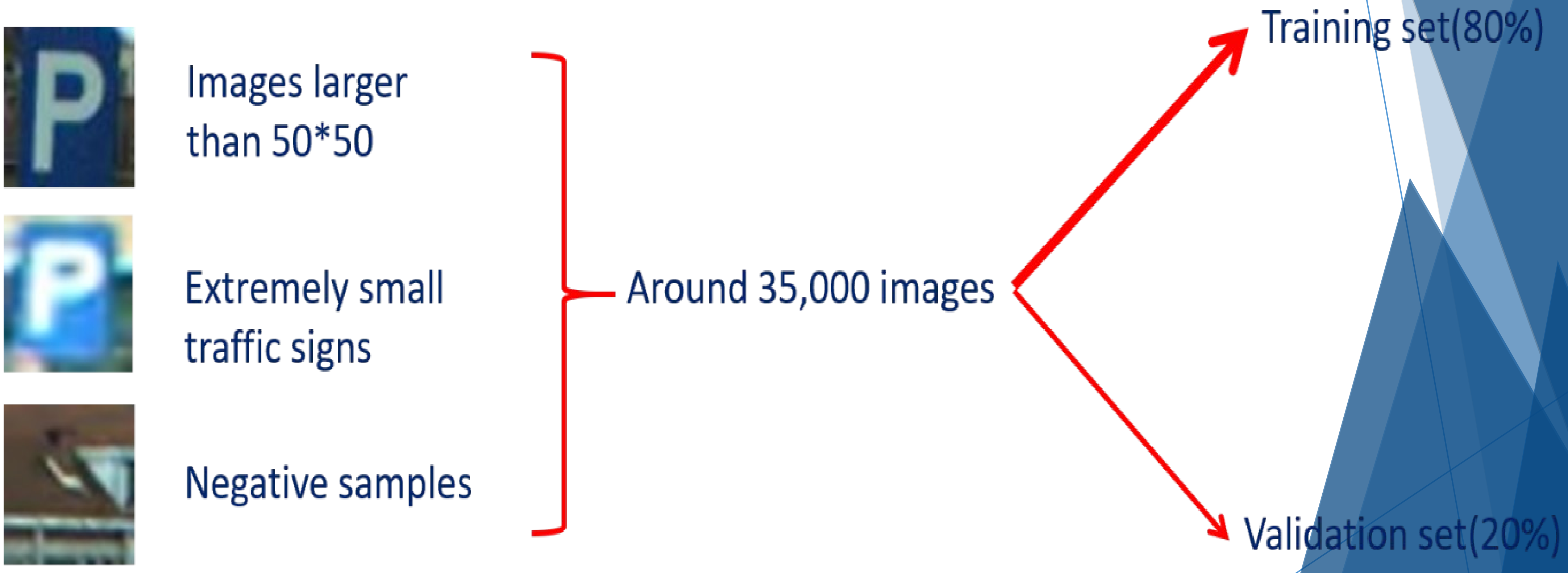
0
0

01 02 03 04 05 06

07 08 09 10 11 12

00-no challenge	07-gaussian blur
01-decolorization	08-noise
02-lens blur	09-rain
03-codec error	10-shadow
04-darkening	11-snow
05-dirty lens	12-haze
06-exposure	

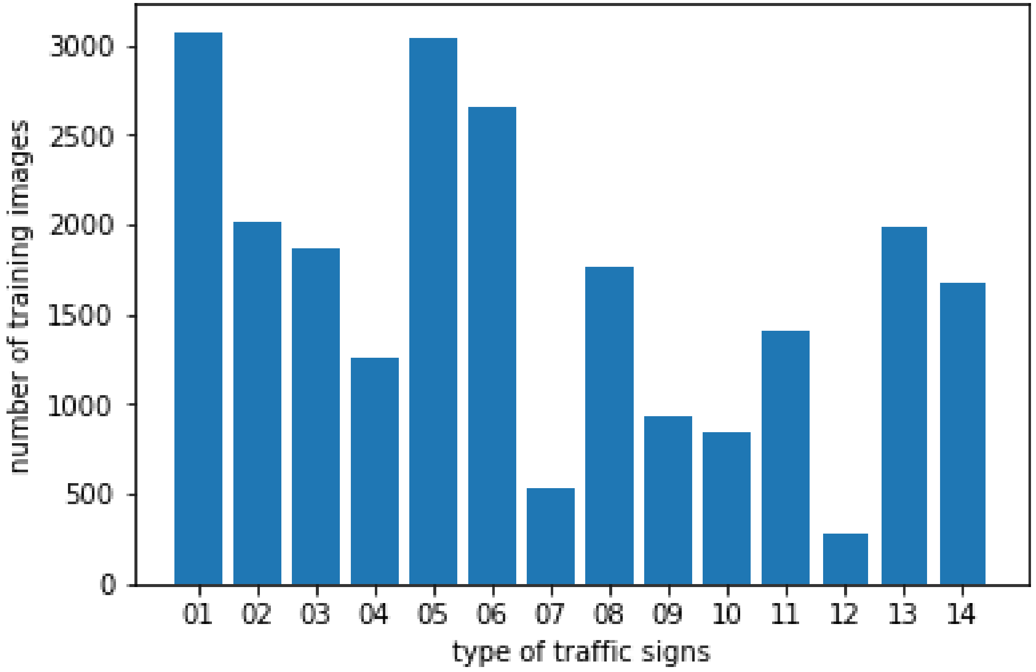
Data preparation



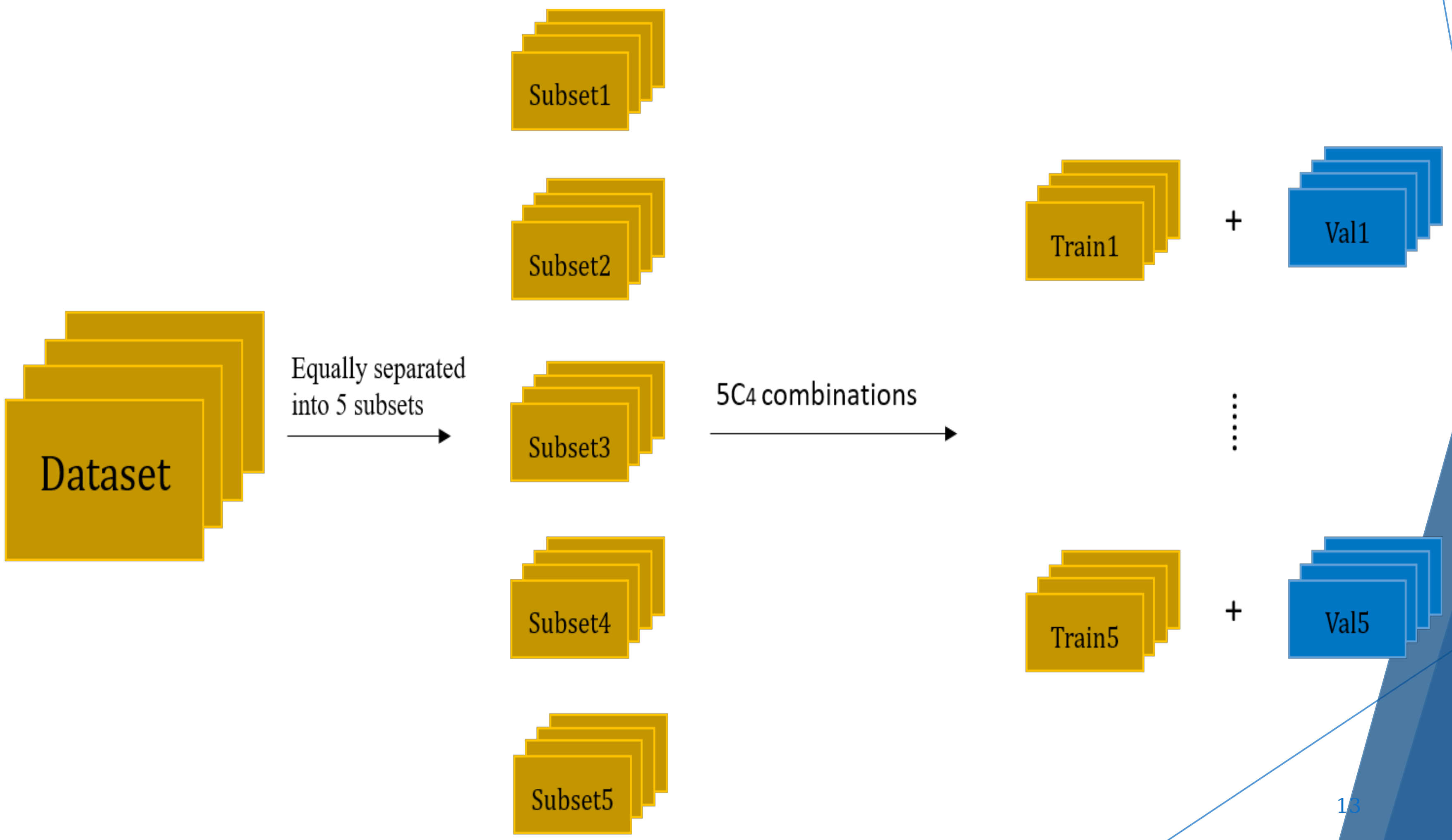
Data augmentation

- Compensation on type 07 and 12

- Random crop
- Random saturation
- Random contrast
- Vertical flip



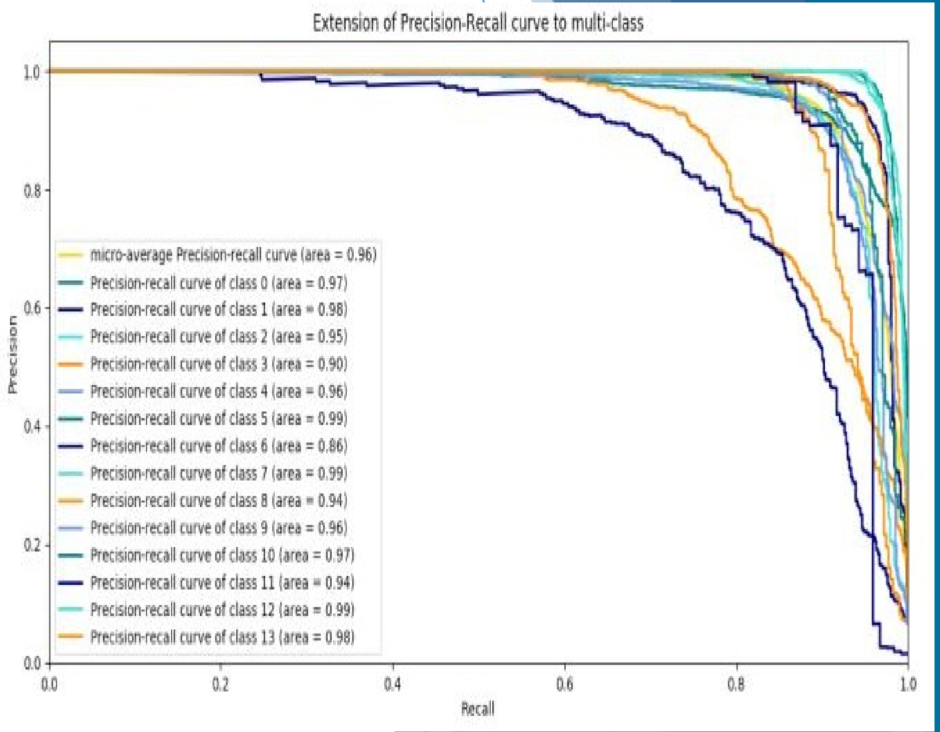
Data preparation



Experimental results

- Running our trained network on 12159 testing images. We gained an average accuracy 92.21%

Challenge Type	Accuracy(%)	Challenge Type	Accuracy(%)
No challenge	99.19	Gaussian blur	91.41
Decolorization	96.99	Noise	96.73
Lens blur	94.07	Rain	92.27
Codec error	59.71	Shadow	97.59
Darkening	96.48	Snow	95.62
Dirty lens	98.36	Haze	95.19
Exposure	93.04	Overall	92.21



- We found that codec error (challenge 03) is the most influential condition on our network. Around 47% of the misclassified images belong to the codec-error challenge, among which 74% are higher than level 2. Figure 2 shows some images of the codec-error challenge, and some images belonging to other challenges.

Experimental results

In order to show the effect of batch normalization and ReLU in this experiment. We tried different combinations of activation function and batch normalization. The experiment results are shown below.

Combination	**ReLU only**	**tanh + BN**	**ReLU + BN**
Accuracy	77.53%	83.38%	92.21%

Experimental results

Dataset	Including negative samples		
Network	Single CNN	Ensemble – Averaging	Ensemble – linear combination
Accuracy	92.21%	92.94%	93.13%

Dataset	Excluding negative samples		
Network	Single CNN	Ensemble – Averaging	Ensemble – linear combination
Accuracy	94.10%	95.17%	95.19%

Future works

- Attempt more pre-processing methods

Some pre-processing techniques are quite useful in traffic sign recognition, such as *localized histogram equalization*, *Gamma correction etc.*

It is also suggested in [3] that using color images does not have any difference from using grayscale images. Therefore, we may simple use grayscale images.

Besides, I would like to try if the convolutional neural network could learn the challenging conditions. If it succeeds, I would like to have the ConvNets classify the challenging type first and then do classification based on the challenging type

- Other models…

Multi-scale CNN, Inception, ResNet…

Future works

- Hand-crafted feature + Random Forest

References

[1] J. Greenhalgh and M. Mirmehdi, “Real-time detection and recognition of road traffic signs,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1498–1506, Dec. 2012.

[2] D. Cireşan, U. Meier, J. Masci and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification", *Neural Networks*, vol. 32, pp. 333-338, 2012.

[3] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale Convolutional Networks," *The 2011 International Joint Conference on Neural Networks*, San Jose, CA, 2011, pp. 2809-2813.

[4] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.

[5] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In ICML, 2013

[6] D. Kingma and J. Ba. Adam: A method for stochastic optimization. ICLR, 2015.

[7] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.