Classification of H&E Stained Breast Cancer Histopathology Images Based on Convolutional Neural Network

Yun Jiang, Li Chen, Hai Zhang and Xiao Xiao

College of Computer Science and Engineering, Northwest Normal University, Lanzhou Gansu, P.R.China

Email: haicheung1995@gmail.com

Abstract. Automatic classification of H&E breast cancer histopathology images is a challenging task. Computer-aided diagnostic systems help reduce costs and increase the efficiency of the process. Although the existing research on breast cancer histopathology image classification is higher than 90% accurate in the binary classifications (non-carcinoma/carcinoma), the classification accuracy of four classifications (normal, benign, in situ, invasive) is less than 80%. This paper proposes a framework for the classification of H&E stained breast cancer histopathological images, which includes two methods based on convolutional neural network. The first method is based on the convolutional neural network structure of the SE-ResNet module, and the second method is based on the transfer learning hybrid model structure, which achieves the accuracy of 80.33% and 86.11% respectively. Compared with the state-of-the-art method, the accuracy is improved by 2.56% and 8.33% respectively. The proposed framework achieves 91.67% accuracy in binary classification and is competitive with state-of-the-art methods.

1. Introduction

It has been reported that the incidence rate of breast cancer ranges from 19.3 per 100,000 women in East Africa, to 89.7 per 100,000 women in Western Europe [1]. Experts can evaluate the microstructure of mammary gland tissue and the type of cancer [2], which provides an important basis for doctors to formulate the best treatment plan [3]. At present, there are some research methods for breast histopathological image classification. FondaN et al. [4] manually constructed a feature vector for each image based on the three sets of features of the kernel, color region and tissue texture of the local and global features of the image, and trained the feature vectors of all images to nine classifiers. The method achieved a four classification accuracy of 75.8%. Zhang et al. [5] proposed a single-class kernel principal component analysis method based on manually design features, with an accuracy of 92%. Araújo et al. [7] enhanced the dataset by randomly collecting image patches from breast cancer histopathological images and used them to train convolutional neural networks, and achieved the four classification accuracy of 77.78% and the binary classification accuracy of 83.33%. Spanhol et al. [8] used the AlexNet network to extract features and classify them using different fusion strategies. The classification accuracy of the binary classifications in the breast cancer pathological image dataset was 6% higher than that of the traditional machine learning algorithm.

It is noteworthy that most of the classification algorithms studied in the current research method use small breast cancer pathological image dataset for training and testing. There have some research difficulties. Firstly, some algorithm based on machine learning algorithm needs a lot of time and energy to extract features manually by using professional domain knowledge. Secondly, which classification algorithm is used to classify the extracted features manually, although, the deep learning avoids manual feature extraction by using the automatic learning method of a deep neural network.
Thirdly, training deep learning network often requires a lot of data to prevent over-fitting and improve the generalization ability of the network. However, histopathological images with labels are often very scarce, and the acquisition process is very expensive.

In this paper, we propose two methods based on convolutional neural networks for classification of histopathological images of breast cancer. In the first method, we construct a new convolutional neural network based on SE-ResNet module, which can automatically extract the features of breast cancer histopathological images and classify them. The second method is to colorize the image with the color normalization data firstly, and then crop out 20 patch images for the enhanced image. Then three different pre-training deep neural networks (GoogLeNet [9], Xception [10], Inception-ResNet [11]) are used to extract the features of the patch image and fuse the patch image of the same image to get the feature vectors. KNN [12] and SVM [12] classifiers were trained to classify breast cancer histopathological images. This method avoids the use of machine learning methods for classification while manually extracting features. According to experiments, the method improves the accuracy, sensitivity, and specificity of histopathological image classification of breast cancer.

2. Automatic Classification Method for Pathological Images of Breast Cancer

2.1. Residual Module

![Residual Module](image)

**Figure 1.** (a) The residual module. (b) Squeeze-and-Excitation module.

Theoretically, with the deepening of the neural network, the capacity and feature extraction ability of the network will continue to increase. However, the common problems of gradient vanishing and gradient exploding of the deep neural network hinder the training of the deep neural network. The residual network proposed by He et al. [13] used the residual module to solve the gradient vanishing problem of deep neural networks, and it won the championship of ILSVRC image classification competition in 2015. The residual module is shown in figure 1 (a). The residual module improves the performance of the network while avoiding the problem of gradient vanishing, which deepens the depth of the network and effectively improves the representation ability of the module. The biggest difference between the residual module and the traditional deep neural networks is that the residual module uses shortcut connections to achieve data overlap between input and output. Assuming that the input and output channels are of the same sizes, let write $x_{i-1}$ for the input of the ith residual module, $f_i(\cdot)$ for multi-layer non-linear mapping, $w_i$ for weight, $b_i$ for deviation, the output of the ith residual module is defined as:

$$y_i \equiv f_i(y_{i-1}, w_i, b_i) + y_{i-1}. \quad (1)$$

The residual module uses two consecutive convolution layers with a convolution kernel size of $3 \times 3$ to convolute the input. The output of the convolutional layer is nonlinearly activated using ReLU, and the data is processed using Batch Normalization. In the backpropagation phase, the parameters are updated using the stochastic gradient descent method, and the gradient formula of the residual module is as follows:

$$\frac{\partial y_i}{\partial y_{i-1}} = \frac{\partial (f_i(y_{i-1}, w_i, b_i) + y_{i-1})}{\partial y_{i-1}} = 1 + \frac{\partial f_i(y_{i-1}, w_i, b_i)}{\partial y_{i-1}}. \quad (2)$$

Equation (2) shows that the gradient of a residual module is always greater than 1 in the process of backpropagation, which solves the problem of gradient vanishing in the process of backpropagation of
deep neural network. A lot of research and experiments also prove that the performance of a deep residual network is better than that of an ordinary deep neural network.

2.2. Squeeze-and-Excitation Block
The Squeeze-and-Excitation Network proposed by Jie Hu et al. [14] used feature recalibration strategy to enable the network to learn the weight of feature channels automatically, to improve network performance and win the championship of ILSVRC image classification competition in 2017. The Squeeze-and-Excitation block (SE block) used in SENet is shown in figure 1 (b). Let \( K = [k_1, k_2, ..., k_{C'}] \) denote a set of convolution kernels, where \( k_{C'} \) refers to the parameters of convolution kernel \( C' \), \( x \) denote the input, and \( O = [o_1, o_2, ..., o_{C'}] \) denote the output, we have \( o_{C'} = k_{C'} \ast x \), where \( \ast \) denotes convolution operation. Since the output is obtained by convolutional summation of the channel and convolution kernel, each channel of the output is interdependent with the convolution kernel, and these dependencies are intertwined with the spatial correlation captured by the kernel. The goal of the SE block is to ensure that the network can increase its sensitivity to features so that the information carried by the features can be leveraged through subsequent transformations and suppress less useful features.

2.3. Model Architecture
The network structure used in this paper is the SE-ResNet network that combines the advantages of the residual module and the Squeeze-and-Excitation module. We propose H&E stained breast cancer histopathology images classification network (HEBCNet), a deep neural network consisting of 1 layer of convolution, nine layers of SE-ResNet module and two layers of the fully connected layer for automatic feature extraction and classification. The entire network is shown in figure 2.

3. Hybrid Model Structure Based on Transfer Learning
Transfer learning is a machine learning algorithm that uses a pre-training model to solve problems in different but related fields. It broadens two basic assumptions in traditional machine learning, with the goal of transfer the existing knowledge to solve the learning problem of only a small number of labeled or unlabeled sample data in the target domain. The training of deep neural networks (such as GoogleLeNet [9], ResNet [13], SENet [14]) often requires a large number of labeled data samples that are very expensive to obtain. In practical applications, few researchers begin to train the whole deep neural network from initialization. The usual method is to pre-train the deep convolution neural network on a very large dataset (such as ImageNet datasets, which contains 1.2 million images of 10,000 categories), and then use the pre-training model as the initialization or feature extractor for the region of interest. In this paper, we use three different pre-trained deep neural networks (GoogLeNet [9], Xception [10], Inception-ResNet [11]) to extract the characteristics of the dataset.

The hybrid structure of transfer learning for the second method proposed in this paper is shown in figure 3. Firstly, we perform 50 randomized staining normalization enhancements for each image to obtain 50 standardized images. Then the image of \( 1024 \times 768 \) pixels is obtained by downsampling,
and then randomly cropping 20 patch images of 512 × 512 pixels from each of the downsampled images. Next, 20 patched images of each image are coded using three kinds of pre-training networks. The sparse features of low dimensions are obtained from the average pooling layer of each network, and 20 feature vectors of each image are obtained. We use the feature fusion method proposed by Bourouei et al. [15] to fuse 20 feature vectors of each image into one feature vector. Then for each original image, we obtain 50 × 3 × 1 = 150 feature vectors. Finally, we train these sparse feature vectors as an input of different classifiers. In this paper, we use the extracted image sparse features and corresponding class labels to train the SVM [12] and KNN [12] and compare their performances.

![Figure 3. The transfer learning hybrid model structure.](image)

4. Experiment and Result

4.1. Dataset and Processing

In this paper, we use the high resolution (2048 × 1536 pixels), uncompressed, labeled and hematoxylin-eosin stained breast cancer histopathological image dataset provided by Araújo et al. [7]. All images were digitized using the same acquisition conditions with a magnification of 200 times and a pixel size of 0.42 μm x 0.42 μm. Each image is labeled as one of four breast cancer types: normal, benign, carcinoma in situ, and invasive cancer. The dataset consists of a training set of 249 images and an independent test set of 36 images. There are 54, 69, 63, and 62 images in the normal, benign, in situ and invasive, respectively. The independent test set is divided into two parts: the first part contains 20 images with the same complexity as the training set, and the second part contains 16 images with higher ambiguity.

Normalization of staining is a crucial step in the processing of hematoxylin-eosin staining images. In this paper, the random normalization method of hematoxylin-eosin staining proposed by Macenko et al. [6] is used. The method first uses a logarithmic transformation to convert the color of the image to an optical density (OD), and then applies singular value decomposition (SVD) to the OD tuple to find a 2D projection with a higher variance. The resulting color space transform is then applied to the original image, and the histogram equalization is performed on the image finally. In this paper, the dataset is expanded by 50 times by performing 50 randomizations for each image in the dataset, and then an image of 224 × 224 pixels is obtained by downsampling as an input to the network.

4.2. Training Strategy

In this paper, the experimental hardware configuration: Intel I7 CPU, NVIDIA 1060Ti GPU (6G). In this paper, we use python as a programming language on Ubuntu 16.04 system and use the deep learning framework of Keras and Sklear. A hierarchical k-fold cross-validation method is used to generate five folds randomly, by keeping the percentage of each folded subset equaling to that of the original data class samples. Each fold contains 198 training sets and 51 validation sets. To ensure the accuracy of the experiment, the randomized normalized extended image and the extracted feature vector of each original image must be contained in the same subset.
For the HEBCNet, we use the stochastic gradient descent method to update the network parameters. We set momentum = 0.9, weight decay=0.0001, the small batch size = 20, the training period = 300, and the initial learning rate = 0.1. The network is initialized by the method proposed by Kaiming He et al. [13]. In the training process, the input image is subjected to random 40-degree rotation, random horizontal and vertical offset with 0.125 amplitude, sheer transformation at 0.2, random zoom amplitude of 0.2, random horizontal flip and other data augmentation, which prevents model overfitting and improves the generalization ability of the model.

For the transfer learning hybrid models, it first normalizes each original image and crops 20 patch images. Then, 20 feature vectors are extracted from a patched image using pre-training model, and a feature vector is obtained by fusing 20 feature vectors using 3-Norm pooling method [16]. Finally, the obtained feature vectors are used to train the KNN and SVM classifiers. After training each model, the final test is performed on each model using an independent test set. The processing method of the independent test set is consistent with the training set, and the category of the pathological image of each piece is voted by the classification result of all the randomized normalized images or feature vectors.

### 4.3. Experimental Results

The classification results of the model proposed in this paper are shown in Table 1. According to the results shown in Table 1, the average classification accuracy of HEBCNet is 82.35% in each fold, which is better than most of the other classification models. For the transfer learning hybrid models, the classification accuracy of KNN GoogLeNet is 78.43%. The classification accuracy of SVM Xception is 82.75% and better than KNN and HEBCNet. According to Table 1, we can calculate the average classification accuracy of all models on five folds to reach 78.60%. In [4], a 5-fold cross-validation method was applied to the dataset. FondaN et al. constructed a feature vector for each image manually and trained nine classifiers (including KNN classifiers with different distance functions and SVM classifiers with different kernel functions) based on three sets of features related to local and global features, color regions and textures of the image. The KNN classifier with cubic polynomial distance achieved a classification accuracy of 75.0%, and the quadratic polynomial SVM classifier achieved a classification accuracy of 75.8%. According to the results in Table 1, the average classification accuracy of KNN GoogLeNet, SVM Xception and HEBCNet is 2.62%, 6.55% and 6.95% higher than FondaN et al. The methods in this paper does not need to construct feature vectors for each image manually. It not only reduces the complexity of manual processing but also proves the effectiveness of the convolutional neural network structure and the hybrid model based on transfer learning. We also find that features extracted automatically by neural networks are more helpful for image classification.

| Model                | Fold 1(%) | Fold 2(%) | Fold 3(%) | Fold 4(%) | Fold 5(%) | Mean(%)  |
|----------------------|-----------|-----------|-----------|-----------|-----------|----------|
| KNN GoogLeNet        | 68.63     | 76.47     | 84.31     | 80.39     | 82.35     | 78.43    |
| KNN Xception         | 62.75     | 68.63     | 82.35     | 78.43     | 80.39     | 74.51    |
| KNN Inception-ResNet | 56.86     | 70.59     | 84.31     | 72.55     | 73.72     |          |
| SVM GoogLeNet        | 72.55     | 70.59     | 86.27     | 92.16     | 80.39     | 80.39    |
| SVM Xception         | 74.51     | 72.55     | 88.24     | 90.20     | 88.24     | 82.75    |
| SVM Inception-ResNet | 68.63     | 74.51     | 84.31     | 76.47     | 86.27     | 78.04    |
| HEBCNet              | 80.39     | 74.51     | 92.16     | 82.35     | 82.35     | 82.35    |

Comparing the average classification accuracy of each fold, we can know that the classification accuracy of each model in the fold 1 and fold 2 is lower than other folds. Our analysis is as follows: because of the subtle differences in the production process, cell overlap, uneven color distribution, and other factors, some of the images are more difficult to classify. In the data preprocessing process, 20
patch images are obtained by randomly cropping each image and used to extract features. Since the abnormal location information of histopathological images is not used in the process of random clipping, new noise will be introduced into the random clipping operation, which will lead to the reduction of classification accuracy. We believe that the first fold and the second fold of the validation set data may contain some of the more difficult images or introduce more noise during random processing to reduce the accuracy.

Table 2. Compare the experiment results with different models on the test set.

| Methods                | Four classification (%) | Binary classification (%) |
|------------------------|--------------------------|----------------------------|
|                        | First part | Second part | All   | First part | Second part | All   |
| Quadratic SVM          | 75.00      | 61.11       | -     | -          | -          | -     |
| CNN Majority           | 80.00      | 75.00       | 77.78 | 80.00      | 81.25      | 80.56 |
| CNN SVM Majority       | 85.00      | 68.75       | 77.78 | 90.00      | 75.00      | 83.33 |
| KNN GoogLeNet          | 85.00      | 68.75       | 77.78 | 85.00      | 75.00      | 80.56 |
| KNN Xception           | 85.00      | 62.50       | 75.00 | 90.00      | 75.00      | 83.33 |
| KNN Inception-ResNet   | 75.00      | 75.00       | 75.00 | 85.00      | 81.25      | 83.33 |
| SVM GoogLeNet          | 85.00      | 81.25       | 83.33 | **95.00**  | **87.50**  | **91.67** |
| SVM Xception           | 80.00      | 75.00       | 77.78 | **95.00**  | 81.25      | 88.89 |
| SVM Inception-ResNet   | 85.00      | **87.50**   | **86.11** | 90.00      | 87.50      | 88.89 |
| HEBCNet                | 85.00      | 75.00       | 80.56 | **95.00**  | 81.25      | 88.89 |

Through a detailed analysis of carcinoma categories, we reclassified the data into two categories: non-carcinoma (normal, benign) and carcinoma (in situ, invasive). Four classification tasks and binary classification tasks were performed on all models. The experimental results of the model proposed in this paper on the independent training set and the accuracy of the baseline are shown in table 2. According to table 2, we can conclude that for the first part of the independent test set, the four classification accuracy and the binary classification accuracy of the model are 85% and 95.00% respectively, which are better than the experimental results in [4] and [7]. Since the second part of the independent test set has higher ambiguity and stronger noise, it has higher requirements for the generalization of the model. The four classification accuracy and binary classification accuracy of the model in the second part of the independent test set are 87.50%, which are 12.50% and 6.25% higher than the experimental results of CNN Majority method in [7]. By fusing the classification results of the first and second parts of the independent test set, the four classification accuracy of SVM Inception-ResNet, SVM GoogLeNet, and HEBCNet are 86.11%, 83.33% and 80.56% respectively, which are higher than the results of [4] and [7].

The results confusion matrix, sensitivity, and specificity of the SVM Inception-ResNet model for independent test sets are shown in table 3. Table 3 shows that the average sensitivity and specificity of SVM Inception-ResNet model are 86.11% and 86.88% respectively. In the binary classification results of the independent test set, the accuracy of the proposed SVM GoogLeNet model is 91.67%, which is 8.34% higher than the experimental results of CNN SVM Majority method [7]. The proposed models of SVM Xception, SVM Inception-ResNet and HEBCNet have achieved 88.89% accuracy of classification, which is 5.56% higher than the CNN SVM Majority method proposed by Araújo et al. [7]. The prediction results confusion matrix, various sensitivity and specificity of the SVM GoogLeNet model for independent test sets are shown in table 4. Table 4 shows that the average sensitivity and average specificity of the SVM GoogLeNet model in the binary classification are 91.67% and 92.86%, respectively. Although SVM Inception-ResNet achieves the highest accuracy among the four classifications, it can be inferred from table 3 that the binary classifications of four data in the model are wrong and the binary classifications accuracy is 88.89%.

---

6
Table 3. The confusion matrix, sensitivity and specificity of SVM Inception-ResNet model.

|                  | Non-carcinoma | Carcinoma |
|------------------|---------------|-----------|
|                  | Normal        | Benign    | In situ   | Invasive |
| Non-carcinoma    | 9             | 0         | 0         | 0        |
|                  | 1             | 7         | 1         | 0        |
| Carcinoma        | 0             | 2         | 7         | 0        |
|                  | 0             | 1         | 0         | 8        |
| Four classification sensitivity (%) | 100.0 | 77.78 | 77.78 | 88.89 |
| Four classification specificity (%)  | 90.00 | 70.00 | 87.50 | 100.0 |
| Binary classification sensitivity (%) | 94.44 |         | 83.33 |         |
| Binary classification specificity (%)  | 85.00 |         |         | 93.75 |

Table 4. The confusion matrix, sensitivity and specificity of SVM GoogLeNet model.

|                  | Non-carcinoma | Carcinoma |
|------------------|---------------|-----------|
|                  | Normal        | Benign    | In situ   | Invasive |
| Non-carcinoma    | 7             | 2         | 0         | 0        |
|                  | 1             | 8         | 0         | 0        |
| Carcinoma        | 0             | 2         | 7         | 0        |
|                  | 0             | 1         | 0         | 8        |
| Four classification sensitivity (%) | 77.78 | 88.89 | 77.78 | 88.89 |
| Four classification specificity (%)  | 87.50 | 61.54 | 100.0 | 100.0 |
| Binary classification sensitivity (%) | 100.0 |         | 83.33 |         |
| Binary classification specificity (%)  | 85.71 |         |         | 100.0 |

In conclusion, the experimental results and analysis show that the proposed framework for breast cancer histopathological image classification achieves higher classification accuracy, sensitivity, and specificity. Compared with the experimental results of Araújo et al. and FondaN et al., the binary classification schemes for breast cancer pathology images have improved the classification accuracy of breast cancer histopathological images. Compared with the convolutional neural network that was constructed by Araújo et al., we construct the module based on convolutional neural network structure in this paper has better classification results in four classifications and binary classifications on breast cancer histopathological images. Compared with the method of Araújo et al. and FondaN et al., the second model based on transfer learning constructed in this paper achieves higher classification accuracy, which proves the effectiveness of the hybrid model structure of transfer learning.

5. Conclusion

In this paper, we propose two methods based on convolutional neural network for classification the H&E stained breast cancer histopathological images. The convolutional neural network of the first method shows that the HEBCNet achieves higher classification accuracy on the breast cancer histopathological image dataset. The second method is based on the transfer learning hybrid model structure. By using different pre-training models to extract features and training two classifiers, the complexity and limitation of manual extraction of features are avoided, and the classification accuracy is improved. Experiments show that the classification accuracy of the transfer learning hybrid model structure is much higher than that of the manual extraction feature in [4] and [7]. The framework of breast cancer histopathological image classification proposed in this paper has robustness and generalization ability. The method also meets the higher requirements of computer-aided diagnosis systems to a certain extent.
6. Acknowledgements

The relevant funds of this paper were supported by: (1) The National Science Foundation Funding Project (61163036). (2) Program of NSFC Financing for Natural Science Fund in 2016. (3) The Special Fund Project of Basic Scientific Research Operating Expenses of Institutes and Universities in Gansu in 2012. (4) Institutes and Universities Graduate Tutor Project in Gansu (1201-16). (5) The Third Period of the Key Scientific Research Project of Knowledge and Innovation Engineering of the Northwest Normal University (nwnu-kjcxgc-03-67).

7. References

[1] Rahhal A and Mahmoud M 2018 Breast cancer classification in histopathological images using convolutional neural network *International Journal of Advanced Computer Science and Applications* 9(3) pp 64-68.

[2] Kowal M, Filipczuk P, Obuchowicz A, et al. 2013 Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images *Computers in Biology and Medicine* 43(10) pp 1563-72.

[3] Yousefi M, Krzyżak A and Suen CY 2018 Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning *Computers in Biology and Medicine* 96 pp 283-293.

[4] Fonda N, Sarmiento A, Garca-A A I, et al. 2018 Automatic classification of tissue malignancy for breast carcinoma diagnosis *Computers in Biology & Medicine* 96 pp 41-51.

[5] Zhang Y, Zhang B, Coenen F, et al. 2014 One-class kernel subspace ensemble for medical image classification *EURASIP Journal on Advances in Signal Processing* 2014(1):17.

[6] Wang P, Hu X, Li Y, et al. 2016 Automatic cell nuclei segmentation and classification of breast cancer histopathology images *Signal Processing* 122 pp 1-13.

[7] Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A and Campilho A. Classification of breast cancer histology images using convolutional neural networks. *PLOS ONE*. 12(6), e0177544.

[8] Macenko M, Niethammer M, Marron J S, et al. 2009 A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro.* IEEE pp 1107-10.

[9] Spanhol F A, Oliveira L S, Petitjean C and Heutte L 2016 Breast cancer histopathological image classification using Convolutional Neural Networks In *2016 International Joint Conference on Neural Networks (IJCNN).* IEEE pp 2560-67.

[10] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D and Rabinovich A 2015 Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pp 1-9.

[11] Chollet F. 2017 Xception: Deep learning with depthwise separable convolutions preprint *Preprint* arXiv: 1610.02357.

[12] Szegedy C, Ioffe S, Vanhoucke V and Alemi A A 2017 Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence.*

[13] Benmahamed Y, Teguar M and Boubakeur A 2017 Application of SVM and KNN to Duval Pentagon 1 for transformer oil diagnosis *IEEE Transactions on Dielectrics and Electrical Insulation* 24(6) pp 3443-51.

[14] He K, Zhang X, Ren S and Sun J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* pp 770-78.

[15] Hu J, Shen L and Sun G 2017 Squeeze-and-excitation networks *Preprint* arXiv: 1709.01507.

[16] Bourau Y L, Ponce J and LeCun Y 2010 A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)* pp 111-118.