EEEM066
Deep Convolutional Neural Networks for Fine-Grained Knife Classification

A disciplined approach to transfer learning

Elena Mylona, URN: xxx, EM02069@surrey.ac.uk

Abstract

The global increase in knife-relate crimes, especially in England and Wales where offences rose by 9% in the year ending March 2022, underscores the necessity for a deep learning weapon analysis. Thus, we propose …
In this study, the composed parts of transfer learning approach, pre-training and fine-tuning hyperparameters of neural networks, has been applied to the above formulated problem of weapons achieving optimal accuracy for our multi-class knife classification system.
Ablation studies
Related work ..to further refine the accuracy and efficiency of the models…
Fine grained object recognition powered by deep learning.
In our experiments we show the results that EfficientNet, DenseNet and DeiT pre-trained models …
We scrutinised our experimental results and analysed …
Peak accuracy and findings …. With the best deep ConvNet …

1. Introduction

Utilised deep CNNs which were prior trained on the ImageNet dataset, EfficientNet-B0, EfficientNet-B6 as well as DenseNets and the DeiT transformer. In our approach, we adopt the common procedure of transfer learning, emphasizing on fine-tuning hyperparameters to achieve optimal accuracy of the model which well suits our formulated problem of a multi-class knife image classification system.

Discarded the last fully connected layers for classification of the different pre-trained models that are being used for these experiments. Added new layers for the specified 192 classes of knives, and then re-trained those new convolutional layers along fine-tuned their hyper-parameters. The training dataset consists of around 9928 knife images in 192 classes. The test dataset consists of 351 images. The training, validation and test splits are provided. The performance of the model will be evaluated using mean average precision (mAP).

Intra-class variance to combat class imbalance trough advanced data augmentations methods and inter-class discrepancy via DL models with modified layers which learn high-level features …

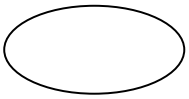

Figure 1: Example of a figure with caption. Captions are set in roman, 9 point. Use a Drawing

1.1. Deep CNNs Model Architectures

All manuscripts must be in English. **CHANGE**
Training strategies have been applied to .. learn a discriminative CNN for representation learning.
The primary challenge in training a deep ConvNets was linked to the scheduling of the learning rate and the degree of L2 weight decay regularisation invoked.

Attempts to freeze some layers of all neural networks..

1.1.2 Dataset Split Usage

Fwwrwrwrwrwrwrw
Dd
Mention the corresponding modes for train, val, and test.

Prepared the validation and test data by preprocessing and not augmenting our knife images. It involved keeping the images to a consistent size, converting them to tensors, and normalizing their pixel values as it aligns the input data distribution with the distribution of the data images used during the pretraining of the models on ImageNet dataset. Hence, the input data remains in the correct format and scale for the models but do not introduce the variability or new samples characteristic of data augmentation since it does not randomly alter or expand the dataset.

1.1.3 Samples of Images with and without Noise

(Include their integer label and class names)

1.2. Hyper-Parameters

Experimented with different settings by fine-tuning hyper-parameters such as the learning rate, batch size optimizers, weight decay and momentum.

The learning rate $\boldsymbol{\eta}$ is a hyperparameter that determines the step size during the gradient descent optimization process. It controls how much the weights in the neural network are updated during training. The slide depicts three scenarios:

Small Learning Rate: Progress is slow, and it takes a long time to converge to the minimum of the loss function.

Large Learning Rate: The updates may be too large, causing the optimization to overshoot the minimum, leading to possible divergence.
Optimal Learning Rate: Achieves efficient convergence to the minimum, balancing the size of steps to neither be too small nor too large.
If the learning rate is changed, you might need to adjust:
Batch Size: A smaller learning rate might benefit from a larger batch size and vice versa.
Number of Epochs: With a smaller learning rate, you might need more epochs to converge.
Learning Rate Schedule: Implementing a schedule that adapts the learning rate during training can help with convergence.
These hyperparameters are interdependent, and changes to one can necessitate adjustments to others.

1.2.1 Cosine Annealing Scheduler

1.2.2 Weight Decay

Uruuyru fsf sfsjfbsjfbsdfs sddddddddddddddddddd dddf fffff

https://paperswithcode.com/method/weight-decay

1.2.3 Momentum

Enhancement over stochastic gradient descent (SGD) – momentum helped accelerate SGD in the relevant direction and dampens oscillations.

1.2.4 Adaptive Moment Estimation - Adam

Optimizers
Adam
AdamW
SGD optimizer

Insert graphs

1.3. Ablation Studies

To effectively train deep models
Perform backpropagation/ backpropagate …

- Weights remove from loss function

-Models experimentation (discuss diff model architectures tested (number of layers) & their performances)
- Justify the choice of the final model architecture based on empirical results.

The experimental results from our knife multi-class image classification robustly corroborate the efficacy of our proposed DCNN, demonstrating its suitability in fine-grained knife classification powered by deep ConvNets.

$$E = m \cdot c^2 \quad (1)$$

and

$$v = a \cdot t. \quad (2)$$

Model Calibration: ECE & Brier Score

1.3.1 Loss Functions Alignment

Cross Entropy for multi-class classification
Weighted Loss function
LF without weights

Impact on **Model Convergence**

1.4. The Significance of Random Seed Setting

This is for reproducible results….
Consistent in neural network training

1.5. Data Augmentation

Refers to the process of applying random transformations to training data to generate new and varied samples, which helps in preventing overfitting and improving models' generalisation capabilities.

Since some classes consist of fewer knife images

1.6. Handling Class Imbalance

Weighted loss,

1.7. Custom-Designed Neural Network

Batch normalization applied to the layers…
Implemented a Depthwise convolutional layer
Squeeze and Excitation Network introduced a building block
https://towardsdatascience.com/squeeze-and-excitation-networks-9ef5e71eacd7

2. The Proposed Knife Classification System

Based on our empirical results, the optimal model architecture … which is suggested to

Ffs

https://arxiv.org/pdf/1708.03211.pdf

3. Experimental Analysis of Results

All text must be in a two-column format. The total allowable size of the text wide.

Corroborate

3.1. Implicit Regularisation – Dropout Rate

Removing layers – technique to prevent overfitting.

https://aws.amazon.com/what-is/overfitting/

Model A achieved a peak accuracy of 98.41% when half of the layers were erased, which substantiates the efficacy of the proposed approach. Model A with dropout rate of 0.5 has less number of neurons as compared to Model B with 0.7 dropout rate in fully connected layers.

3.2. Label Smoothing

Label smoothing cross entropy

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title $1\frac{3}{8}$ inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

2.2.1. Additive Angular Margin Loss (ArcFace)

Although, this method is primarily used for face recognition, it also holds potential for complex multi-class classification tasks. The rationale for employing this loss function in our experimentation within deep ConvNets is to maximise the angular margin between the correct class and the remaining classes of knives.

In ArcFace, the core concept involves shifting the similarity measure from Euclidean distance, L2 Norm, to Cosine distance, primarily due to its invariance to partial noise within the data.

https://shubham-shinde.github.io/blogs/arcface/

insert diagram & EQUATION

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1 Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

3.3. Boosting Ensemble Mechanisms

Boosting here

4. Evaluation Metrics

4.1.1. Accuracy, Precision, Recall

maP during training and then testing

Observation 1: ddsdsds

Observation 2: gerg

Precision – Specificity:

Recall – Sensitivity:

Gradual reduction in validation loss

Insert plot (y-accuracy / x-training time)

Top-1 Accuracy

Top-5 Accuracy

Training and Validation Loss Calculation & Plot

Training and Validation Accuracy Plot (Top-1 and Top-

Validation mAP Plot

4.1.1 F1 score

Harmonic …

4.1.2 Test and Validation Errors

Dddddd

fddsa

4.1.3 Model Convergence and Divergence

Gdgd

fff

4.2. Macro Average

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [5]. Where appropriate, include page numbers and the name(s) of editors of referenced books. When you cite multiple papers at once, please make sure that you cite them in numerical order like this [1, 2, 4-6].

4.3. Weighted Average

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Readers (and reviewers), even of an electronic copy, may choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

4.4. Multi-Class Confusion Matrix

If you use color in your plots, please keep in mind that a significant subset of reviewers and readers may have a color vision deficiency; red-green blindness is the most frequent kind. Hence avoid relying only on color as the discriminative feature in plots (such as red *vs.* green lines) but add a second discriminative feature to ease disambiguation.

4.4.1. Weakness of Samples Misclassification

Ddd

ww

4.5. Probability Calibration in Deep CNNs

Considering the prowess of deep convolutional networks in object-recognition tasks from images and yield probabilities for each class in classification operations, we directed our attention towards on calibration curves. We assessed how well the predicted probabilities made by these neural networks align with the actual outcomes.

+ +

Insert plot

```
from sklearn.calibration import
calibration_curve,
CalibratedClassifierCV
```

4.6. DenseNets and Heatmaps

Gdgfd

Fdfd

5. Future Directions

Our future endeavors will be centered around model convolutional neural network scaling for comparable speedups on memory-limited hardware, i.e., GPU, TPU. Moreover, TTA …

5.1. Scaling up models for ConvNets

An approach is to scale a Baseline Model with Different Network Width (w), Depth (d), and Resolution (r) Coefficients.

" which result in about $O(s)$ increase in model activation w.r.t. scaling flops by a factor of *s*, the proposed fast compound scaling results in close to $O(\sqrt{s})$ increase in activations, while achieving excellent accuracy. "

5.2. Test Time Augmentation (TTA)

Data augmentation during the testing phase holds potential for enhancing the performance of the trained networks. TTA will allow our models to make predictions on multiple augmented versions of each test knife image, thereby increasing their robustness to variations in new data. By averaging the combined predictions from these different versions of images, we anticipate a reduction in prediction errors, particularly in cases where test data differs from training data due to distortions and variations in the test knife images.

5.3. Deep Filters: 5.1.2

Can be viewed as localised descriptors

WEI ET AL.: FINE-GRAINED IMAGE ANALYSIS WITH DEEP LEARNING: A SURVEY

https://arxiv.org/pdf/1708.03211.pdf

in every DL-based classification model, there are two parts inside, namely the feature extraction parts (stacking of convolutional and pooling layers) and classification parts (fully connected layers). When using transfer learning to build a classification model, it is uncertain which parts play a more important role in successfully applying to a new dataset.

In every deep learning-based classification model, two primary components are present: the feature extraction portion, consisting of layers of convolution and pooling, and the classification segment, comprised of fully connected layers. The application of transfer learning in constructing a classification model presents an uncertainty regarding which of these parts – feature extraction or classification – is more critical for effective adaptation to a new dataset.

5.4. Stochastic Gradient Descent with Warm Restarts (SGDR)

In alignment with these findings, it has been proposed in SGDR to employ restart techniques for coordinating the multimodal functions in gradient-free optimization. Accordingly, we could instantiate a warm restart technique for SGD to improve its anytime performance when training DNNs. That is, the partial warm restarts in gradient-based optimisation are being executed in order to enhance the convergence rate in accelerated gradient schemes, specifically for addressing ill-conditioned functions.

Implement the cosine annealing part of SGDR

https://openreview.net/pdf?id=Skq89Scxx (latest)
https://arxiv.org/abs/1608.03983

---

https://shubham-shinde.github.io/blogs/arcface/

https://arxiv.org/abs/2103.06877

References

[1] FirstName Alpher, Frobnication. *IEEE TPAMI*, 12(1):234–778, 2002.

[2] Zagoruyko, S. and Komodakis, N. Wide residual networks. BMVC, 2016

[3] FirstName Alpher and FirstName Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003.

[4] FirstName Alpher, FirstName Fotheringham-Smythe, and FirstName Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.

[5] FirstName Alpher and FirstName Gamow. Can a computer frobnicate? In *CVPR*, pages 234–778, 2005.

[6] FirstName LastName. The frobnicatable foo filter, 2014. Face and Gesture submission ID 324. Supplied as supplemental material `fg324.pdf`.

[7]

[8] https://towardsdatascience.com/squeeze-and-excitation-networks-9ef5e71eacd7

6. Appendix

ArcFace Diagram for Architecture