

Assessment of Deep Convolutional Neural Network Models for Species Identification of Forensically-Important Fly Maggots Based on Images of Posterior Spiracles

Darin Apasrawirote

Naresuan University

Pharinya Boonchai

Naresuan University

Paisarn Muneesawang

Naresuan University

Wannacha Nakhonkam

Naresuan University

Nophawan Bunchu (✉ nophawanb@nu.ac.th)

Naresuan University

Research Article

Keywords: Forensic entomology, Deep Convolutional Neural Network Models, Fly Maggots, Posterior Spiracles

Posted Date: January 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1134033/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on March 19th, 2022. See the published version at <https://doi.org/10.1038/s41598-022-08823-8>.

Abstract

Forensic entomology is the branch of forensic science that is related to using arthropod specimens found in legal issues. Fly maggots are one of crucial pieces of evidence that can be used for estimating post-mortem intervals worldwide. However, the species-level identification of fly maggots is difficult, time consuming, and requires specialized taxonomic training. In this work, a novel method for the identification of different forensically-important fly species is proposed using convolutional neural networks (CNNs). The data used for the experiment were obtained from a digital camera connected to a compound microscope. We compared the performance of four widely used models that vary in complexity of architecture to evaluate tradeoffs in accuracy and speed for species classification including ResNet-101, Densenet161, Vgg19_bn, and AlexNet. In the validation step, all of the studied models provided 100% accuracy for identifying maggots of 4 species including *Chrysomya megacephala* (Diptera: Calliphoridae), *Chrysomya (Achoetandrus) rufifacies* (Diptera: Calliphoridae), *Lucilia cuprina* (Diptera: Calliphoridae), and *Musca domestica* (Diptera: Muscidae) based on images of posterior spiracles. However, AlexNet showed the fastest speed to process the identification model and presented a good balance between performance and speed. Therefore, the AlexNet model was selected for the testing step. The results of the confusion matrix of AlexNet showed that misclassification was found between *C. megacephala* and *C. (Achoetandrus) rufifacies* as well as between *C. megacephala* and *L. cuprina*. No misclassification was found for *M. domestica*. In addition, we created a web-application platform called thefly.ai to help users identify species of fly maggots in their own images using our classification model. The results from this study can be applied to identify further species by using other types of images. This model can also be used in the development of identification features in mobile applications. This study is a crucial step for integrating information from biology and AI-technology to develop a novel platform for use in forensic investigation.

Introduction

The fly larvae of *Chrysomya megacephala* (Diptera: Calliphoridae), *Achoetandrus rufifacies* (Diptera: Calliphoridae), *Lucilia cuprina* (Diptera: Calliphoridae), and *Musca domestica* (Diptera: Muscidae) have long been recognized as crucial clues in medicolegal forensic entomology, especially in Thailand [1–2]. The determination of the minimum time since death (MTD) from fly maggots plays a primary role in estimating the post-mortem interval [3]. Moreover, human myiasis [1], the infestation of fly maggots in a living person, may be found in neglected people. Therefore, fly larvae can be used to estimate the minimum abandonment duration. The accuracy of MTD determination and abandonment duration mainly relies on species identification. The pain point of species identification is the requirement of taxonomic experts, especially in a larval stage. The morphology of larvae is vermiform, circular in cross section, tapering to a point, and having no legs. For the larval stage, the morphology of the posterior spiracle is one of the important characteristics for species identification. Posterior spiracles are the organs involved in the gas exchanges, which are in the last segment and are joined to the dorsal longitudinal trunks of the internal tracheal system. The spiracles are placed in two chitinized plates

surrounded by the peritreme. Typically, the number of slits can be used to classify the stage of larvae. For example, the third-stage larva has three slits in each side of the posterior spiracle. Although the posterior spiracles of some species have specific characteristics, non-taxonomists cannot differentiate them. Species identification based on morphological characteristics is more difficult at the larval stage than at the adult stage. Moreover, it is time consuming, and requires specialized taxonomic training. In addition, the number of taxonomists and classification experts has drastically decreased so far. Although molecular techniques have been accepted for fly species identification, specific equipment and reagents are needed in the laboratory. Because reliability of evidence is the important key for forensic investigation, alternative automatic identification methods with expert-level classification accuracy are highly required in this field. Automatic identification of species is important not only for flies in forensics, but also for other insects in general since it contributes to various purposes such as environment monitoring, pest diagnostics, and vector epidemiology [3]. Many previous studies indicated that automatic classification of insect species achieved high classification accuracy. Currently, many classification models based on convolutional neural networks (CNNs) have been proposed in the field of computer vision. CNN is the type of multi-layered network learning algorithm, which usually contains multiple convolution layers, batch normalization layers, and fully connected layers [4]. Many previous reports demonstrated that the CNNs models have achieved species identification of insects with high accuracy and precision [5–9]. Based on our information, there have been no reports on the identification modeling of forensically-important fly maggots based on CNNs. Therefore, the novel method was proposed and the identification model was developed in 2 parts, including a custom object detection model for detection of position by using Labeling, and YOLO programs, consecutively, and the multiple deep learning models to identify posterior spiracle images of four fly species including *Chrysomya megacephala* (Diptera: Calliphoridae), *Chrysomya (Achoetandrus) rufifacies* (Diptera: Calliphoridae), *Lucilia cuprina* (Diptera: Calliphoridae), and *Musca domestica* (Diptera: Muscidae). All four fly species have been recognized as medically- and forensically-important worldwide [1–2, 10–12]. In this study, the assessment of the identification was compared with state-of-the-art CNNs including ResNet-101, Densenet161, Vgg19_bn, and AlexNet. The best performance model was selected and used to develop a web-application for automatic species identification.

Material And Methods

Preparation of Posterior spiracle

Slides of the posterior spiracle were prepared following the method of Bunchu *et al.* [13]. Briefly, the morphology of the posterior spiracle was investigated by using the hydroxide clearing method. Maggots of each species (*C. megacephala*, *C. (A.) rufifacies* (Diptera: Calliphoridae), *L. cuprina*, and *M. domestica*) were laboratory strains, which were obtained from the Medical Entomology Laboratory, Department of Microbiology and Parasitology, Faculty of Medical Science, Naresuan University. The posterior spiracles of each species were photographed under the light microscope connected to the digital camera. All the images were used in the training process. The identification of all specimens was confirmed by the

experts. For demonstration purposes, example images of each species were inverted using the rgb colors from rgb (255, 255, 255) to rgb (0, 0, 0) by using the image adjustment function of Adobe photoshop 2021 to increase the clarity of the images.

Image data set

In this study, the images of the posterior spiracles of four forensically-important fly species as mentioned above were analyzed. All original images have been categorized as “verified” after they were identified and confirmed by an expert. In total, 17,144 original images were used in this study and divided into training (12,000 images), validation (3,428), and testing (1,714) groups with a number ratio of 70:20:10, respectively. In the first step, we developed a custom object detection model for detection of position of posterior spiracle in the image by using Labelling program. Five hundred images were used for training the object detection model. The custom object detection data set was encoded to the JSON format. After that, this custom data set was implied with the YOLO object detection algorithm to crop images tightly to each species. Cropped images less than 224x224 pixels were discarded from this study. This custom object detection model provided the confidence and accuracy $\geq 80\%$. The process of image detection was shown in Figure 1. The image data set was encoded to the JPEG format for the next model training.

Model training

In this study, we compared the performance of four widely used models that vary in complexity of architecture to evaluate tradeoffs in accuracy and speed for species classification including ResNet-101, Densenet161, Vgg19_bn, and AlexNet. In the first step, we used the model from the previous step as mentioned above to speed up the training process. After randomly shuffling images in each species, the images within the species were divided by 70:20:10 ratio into training, validation, and testing groups. For each model, cropped images were resampled to the standard size for the input model. After using these models initially, we used 224x224 pixel images for Resnet-101, Densenet161, Vgg19_bn, and AlexNet. We used an image augmentation strategy following the previous study by Spiesman *et al.* [14]. Stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.01 for all models were used. We used batch normalization and models were trained for the 50 epochs using Nvidia K80, T4, P4 and P100 GPUs. At the species level, two metric model performance was calculated, including precision and recall. Macro precision or mean of species-level precision scores and macro-recall or mean of species-level recall scores for each model were determined. The mean time needed to predict the image in the test data set was also quantified as the model speed. All speed tests run on the same system using Nvidia K80s, T4s, P4s and P100s GPUs. The best performance model from this study was selected and used to develop web applications for automatic species identification.

Results And Discussion

Of which at the third instar, the external morphology of larvae is quite similar; thus, the morphological identification used to differentiate between its genera or species, generally includes cephalopharyngeal skeleton, anterior spiracle, and posterior spiracles. The morphology of the posterior spiracle is one of the important organs for identification. A typical morphology of the posterior spiracle of third stage larvae was shown in Figure 2. Their morphological characteristics were like the descriptions in the previous reports [15–16]. Based on studying under light microscopy, the posterior spiracle of *M. domestica* was clearly distinguished from the others. On the other hand, the morphology of the posterior spiracle of *C. megacephala* and *A. rufifacies* was quite similar.

For model training, four of the CNN models used for species-level identification of fly maggots provided 100% accuracy rates and 0% loss. Params, model speed, macro precision, macro recall, f-1 score, support value, and model size were also presented in Table 1. As the results in Figure 3 presented, all models provided 100% accuracy and 0% loss in the early stage of training (<10 epochs). This may be due to the training and testing processes with cropping specific portion of the fly images by using our custom object detection model. Moreover, all images were from the laboratory strains, which their variation of morphological characteristics may be less than wild type. Therefore, training time was short, and accuracy of the model was high. Of the four models tested, AlexNet demonstrated a good balance between performance and speed. This model can proceed the system the fastest and its model size is the smallest. The speed and accuracy of AlexNet make it useful for web-based and mobile applications that rely on both speed and reliable predictions. Speed is a factor in user satisfaction and will be important for future development such as video-based applications [14]. Therefore, we focus on the AlexNet results for the remainder of this article.

Table 1

Comparison of model size, speed, and performance of each studied model (The text in bold indicates the best value in each category.)

Model	#Params (million)	Model speed (sec)	Model size (mb)	Macro precision	Macro recall	f1- score	support
Resnet101	44.6	246.0	170.96	1.00	1.00	1.00	3388
Densenet161	14.2	326.4	55.33	1.00	1.00	1.00	3388
Vgg19_bn	20.6	210.0	78.67	1.00	1.00	1.00	3388
AlexNet	2.7	10.8	10.58	1.00	1.00	1.00	3388

The AlexNet model can separate species explicitly into distinct groupings based on characteristics extracted from the model (Figure 4). The results showed that some data of *C. megacephala* overlapped the data of *C. rufifacies*. The visualization of the hidden convolutional layers in four example images of AlexNet is clearly shown in Figure 5.

The classification results (validation and test) for each image are displayed in the confusion matrices (Figure 6) which show how predicted species (columns) correspond to the actual species (rows). The

values along the diagonal indicate the number of correct predictions, whereas off-diagonal values indicate misclassifications. Interestingly, no misclassification was found after testing the model by using the test images (Figure 6A). Therefore, the results indicated that the predictions of AlexNet model match the taxonomic expert classification. To confirm the results of this study, we tested this model with other images from outsources (internet and personal contact) and visualized the results using the PyTorch CNN visualizations [17]. The confusion matrix showed misclassification between *C. megacephala* and *C. rufifacies* (Figure 6B), corresponding to the results of tSNE visualization. When the model was tested with the outsource images, the accuracy of the classification for *C. megacephala*, *C. rufifacies*, *L. cuprina*, and *M. domestica* was 94.94, 98.02, 98.35, 100%, respectively (Figure 6B). The results from using the Heatmap program showed that the prediction accuracy of this model was still high (99.30-100%), depending on image conditions (Figure 7). The Framework of AlexNet model was demonstrated in Figure 8. Previously, CNNs have been used successfully to identify different cells or species [6, 8, 14, 18–19]. This study also confirmed the efficiency of CNNs in identifying fly species.

Finally, we created a web application called “The Fly” by using our classification model for identifying species of fly maggots. The web application is available at <https://thefly.ai>. Users can identify species of fly maggot by uploading their images of posterior spiracles and the result with associated probability percentage will then be shown. This web application can be accessed and used on both desktop and mobile browsers. In terms of performance limitations, this web application was designed to identify only four species of fly maggot using images of posterior spiracles. This web application is the beginning step of the development of automatic species-identification for fly species in Order Diptera. More images of these four species and other species must be studied in the future. In addition, the results from this study will be applied to develop a feature as a microservice for the identification of fly maggots in a mobile application called iParasites which is currently available on AppStore and GooglePlay. We, nonetheless, wish to project that taxonomic experts are still important and critical for the development of this automatic identification by AI-based imaging system as mentioned in a previous report [20].

Conclusion

Currently, CNNs have been successfully used in a wide range of scientific discipline including Entomology. This study demonstrated that all state-of-the-art CNNs used in this study including ResNet-101, Densenet161, Vgg19_bn, and AlexNet provided high efficiency in the computer vision aspect. AlexNet showed the fastest speed to process the identification model and can reliably identify species of fly maggots based on images of posterior spiracles. However, the results from this study confirmed that the AI-based imaging system will be useful in developing automatic species identification for flies and other insects. In the next step, more images of other fly species should be collected as training data sets. Moreover, the results from this study can be used in the development of identification features in mobile applications and in further species identification by using other types of images.

Declarations

Acknowledgments

We are grateful for the financial support from the Naresuan University research fund (R2563C008 to Darlin Apasrawirote and Nophawan Bunchu). Ketsarin Thipphet and Pluemkamon Phuwanatsarunya are acknowledged for their assistance in rearing and maintaining fly colonies in the laboratory.

Author contributions

D.A. conceived the study, contributed model training, and provided acquisition of funding. P.B. performed the modeling, designed all visualization, and created a web application (thefly.ai)., P.M. provided the critical feedback, editing, and approval of the final draft., W.N. prepared slides of posterior spiracles, and took photographs under a stereomicroscope, N.B. initiated and designed the study, identified the specimens, prepared the data sets, performed training of models, wrote the first draft of the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The dataset and source codes for this study are publicly available through the second author's GitHub repository: <https://github.com/pharinyab/Fly-PosteriorSpiracles-AlexNet>

References

1. Bunchu, N. Blow fly (Diptera: Calliphoridae) in Thailand: distribution, morphological identification and medical importance appraisals. *Int. J. Parasitol. Res.* **4**, 57–64 (2012).
2. Sukontason, K., *et al.* Forensic entomology cases in Thailand: a review of cases from 2000 to 2006. *Parasitol. Res.* **101**, 1417–1423 (2007).
3. Harvey, M. L., Gasz, N. E. & Voss, S. C. Entomology-based methods for estimation of postmortem interval. *Res. Rep. Forensic Med. Sci.* **6**, 1–9 (2016).
4. Boukaye, B. T., Bernard, K. F. & Fana, T. Deep convolution neural network for image recognition. *Ecol. Inform.* **48**, 257–268 (2018).
5. Hernández-Serna, A. & Jiménez-Segura, L. F. Automatic identification of species with neural networks. *PeerJ.* **2**, e563; 10.7717/peerj.563 (2014).
6. Motta, D., *et al.* Application of convolutional neural networks for classification of adult mosquitoes in the field. *PLoS One.* **14**, e0210829 (2019).

7. Ong, S. Q., *et al.* Implementation of a deep learning model for automated classification of *Aedes aegypti* (Linnaeus) and *Aedes albopictus* (Skuse) in real time. *Sci. Rep.* **11**, 9908 (2021).
8. Park, J., *et al.* Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Sci. Rep.* **10**, 1012 (2020).
9. Ye, S., Lu, S., Bai, X. & Gu, J. ResNet-Locust-BN Network-Based Automatic Identification of East Asian Migratory Locust Species and Instars from RGB Images. *Insects.* **11**, 458; 10.3390/insects11080458 (2020).
10. Amendt, J., *et al.* Forensic entomology in Germany. *Forensic Sci Int.* **113**, 309–314 (2000).
11. Badenhorst, R. & Villet, M. H. The uses of *Chrysomya megacephala* (Fabricius, 1794) (Diptera: Calliphoridae) in forensic entomology. *Forensic Sci. Res.* **3**, 2–15 (2018).
12. Catts, E. P. & Goff, M. L. Forensic entomology in criminal investigations. *Annu. Rev. Entomol.* **37**, 253–272 (1992).
13. Bunchu, N., *et al.* Morphology and Developmental Rate of the Blow Fly, *Hemipyrellia ligurriens* (Diptera: Calliphoridae): Forensic Entomology Applications. *J. Parasitol. Res.* 2012, 371243. <https://doi.org/10.1155/2012/371243> (2012).
14. Spiesman, B. J., *et al.* Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Sci. Rep.* **11**, 7580 (2021).
15. Sukontason, K. L., *et al.* Larval morphology of *Chrysomya megacephala* (Fabricius) (Diptera: Calliphoridae) using scanning electron microscopy. *J. Vector Ecol.* **28**, 47–52 (2003).
16. Sukontason, K., Piangjai, S., Siri wattanarungsee, S. & Sukontason, K. L. Morphology and developmental rate of blowflies *Chrysomya megacephala* and *Chrysomya rufifacies* in Thailand: Application in forensic entomology. *Parasitol. Res.* **102**, 1207–1216 (2008).
17. Ozbulak, U. Convolutional Neural Network Visualizations. <https://github.com/utkuozbulak/pytorch-cnnvisualizations> (2019).
18. Acevedo, A., *et al.* Recognition of peripheral blood cell images using convolutional neural networks. *Comput Methods Programs Biomed.* **180**, 105020 (2019).
19. Tan, C., *et al.* Identification of different species of *Zanthoxylis Pericarpium* based on convolution neural network. *PLoS One.* **15**, e0230287 (2020).
20. Høye, T. T., *et al.* Deep learning and computer vision will transform entomology. *Proc. Natl. Acad. Sci. USA.* **118**, e2002545117 (2021).

Figures

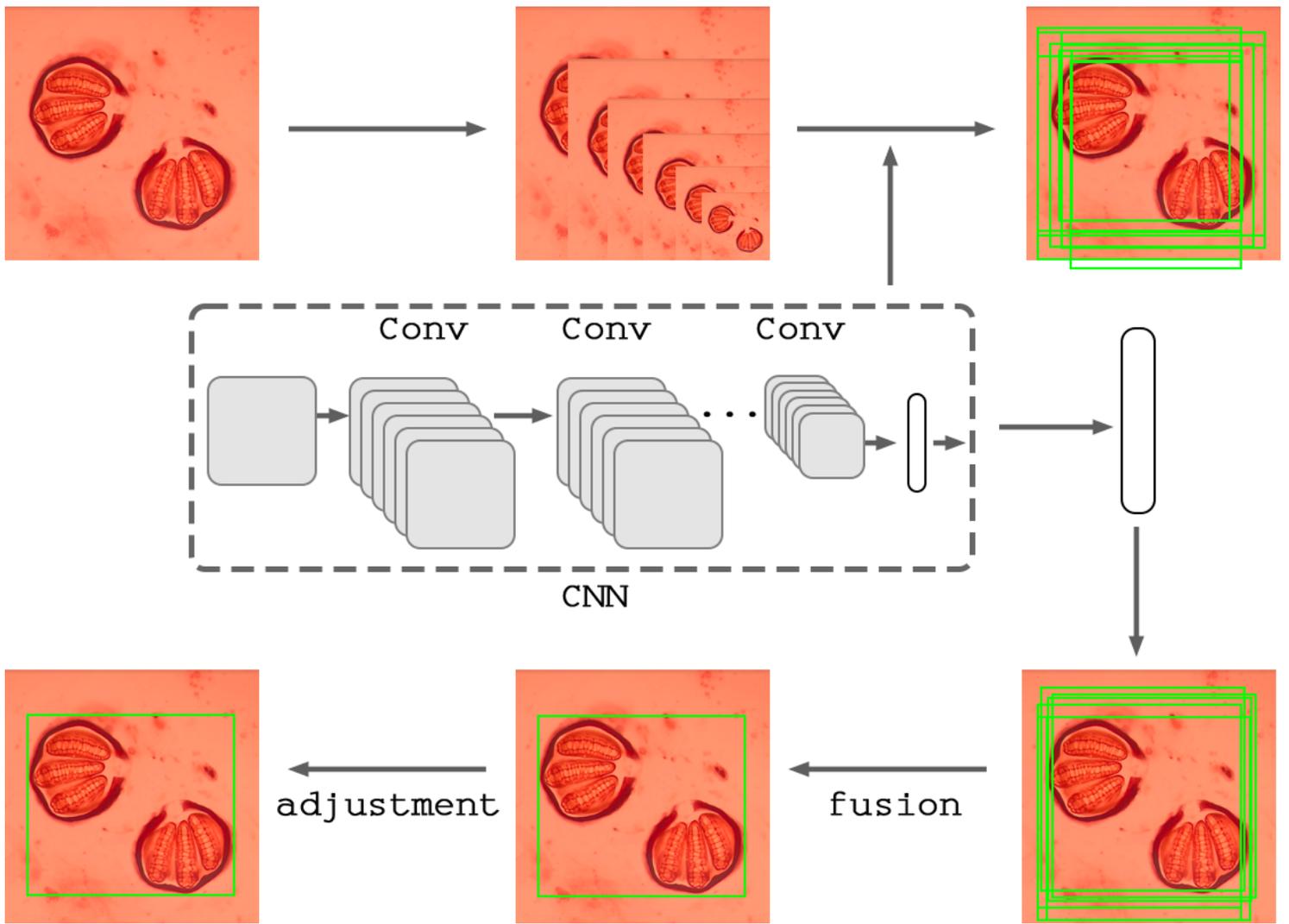


Figure 1

The processes of the custom object detection model used in this study

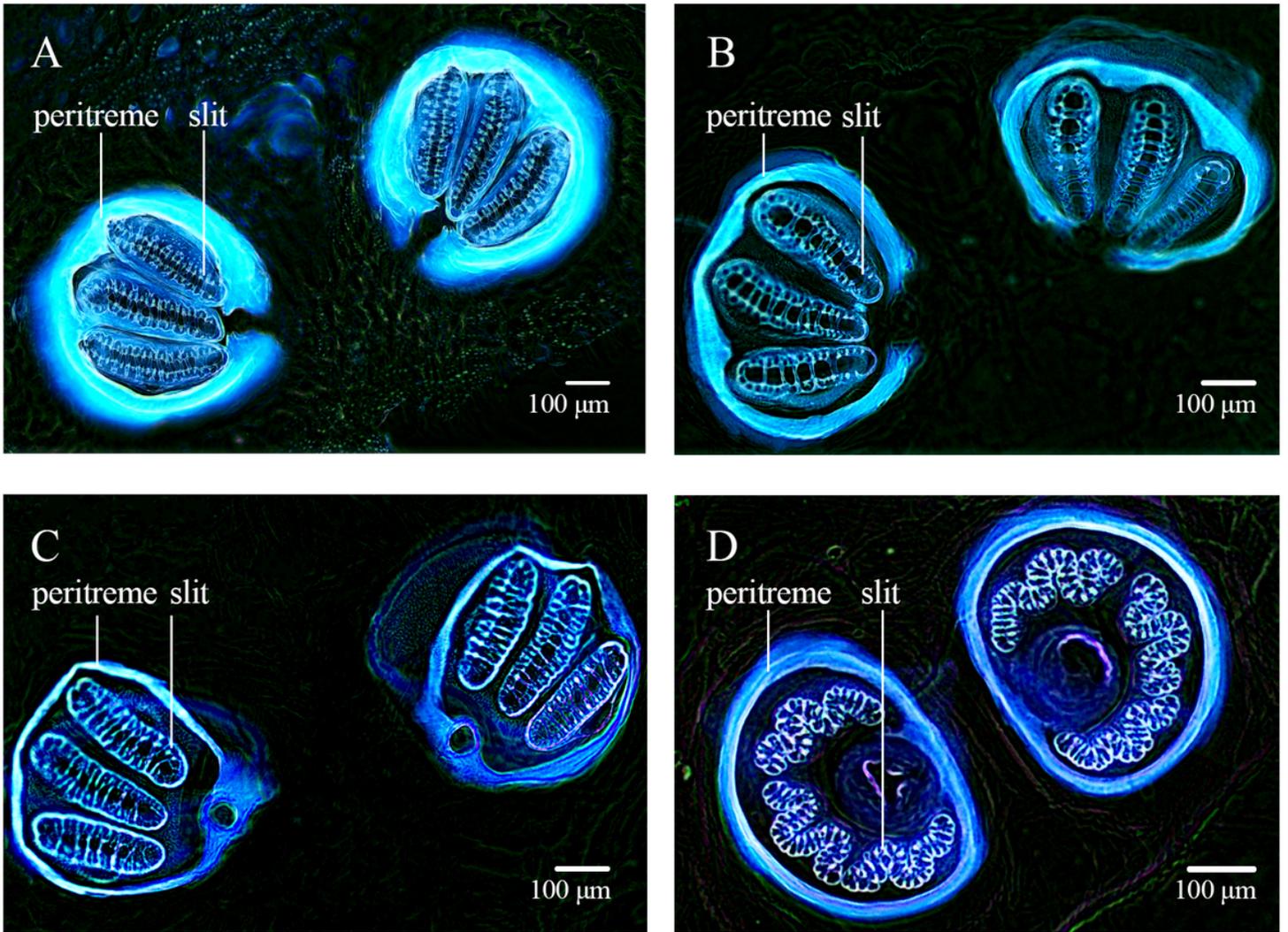


Figure 2

Morphology of posterior spiracles of four different fly species after inverting the image colors; A: *Chrysomya (Achoetandrus) ruffifacies*, B: *Chrysomya megacephala*, C: *Lucilia cuprina*, D: *Musca domestica*

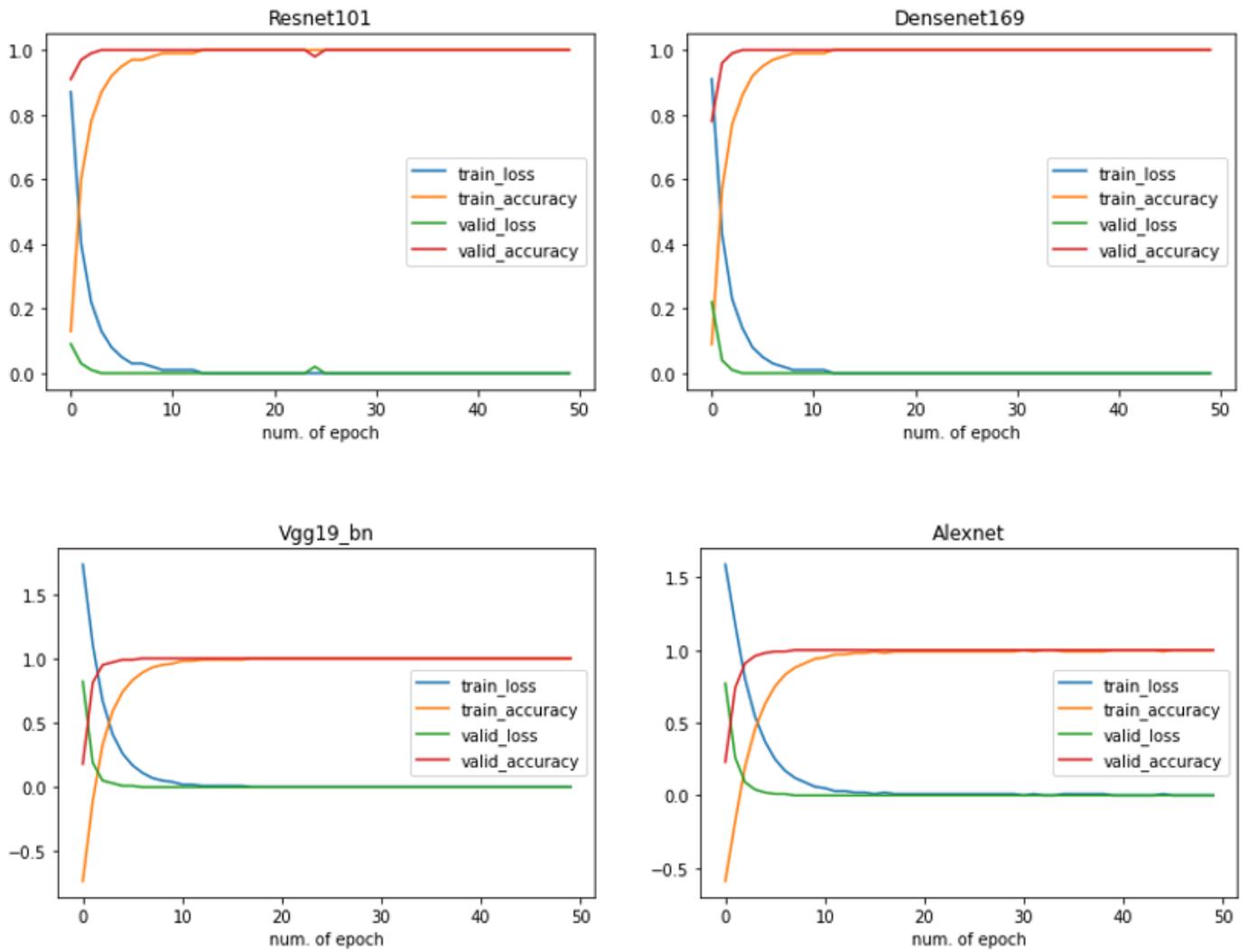


Figure 3

Validation accuracy of the four different models during the training

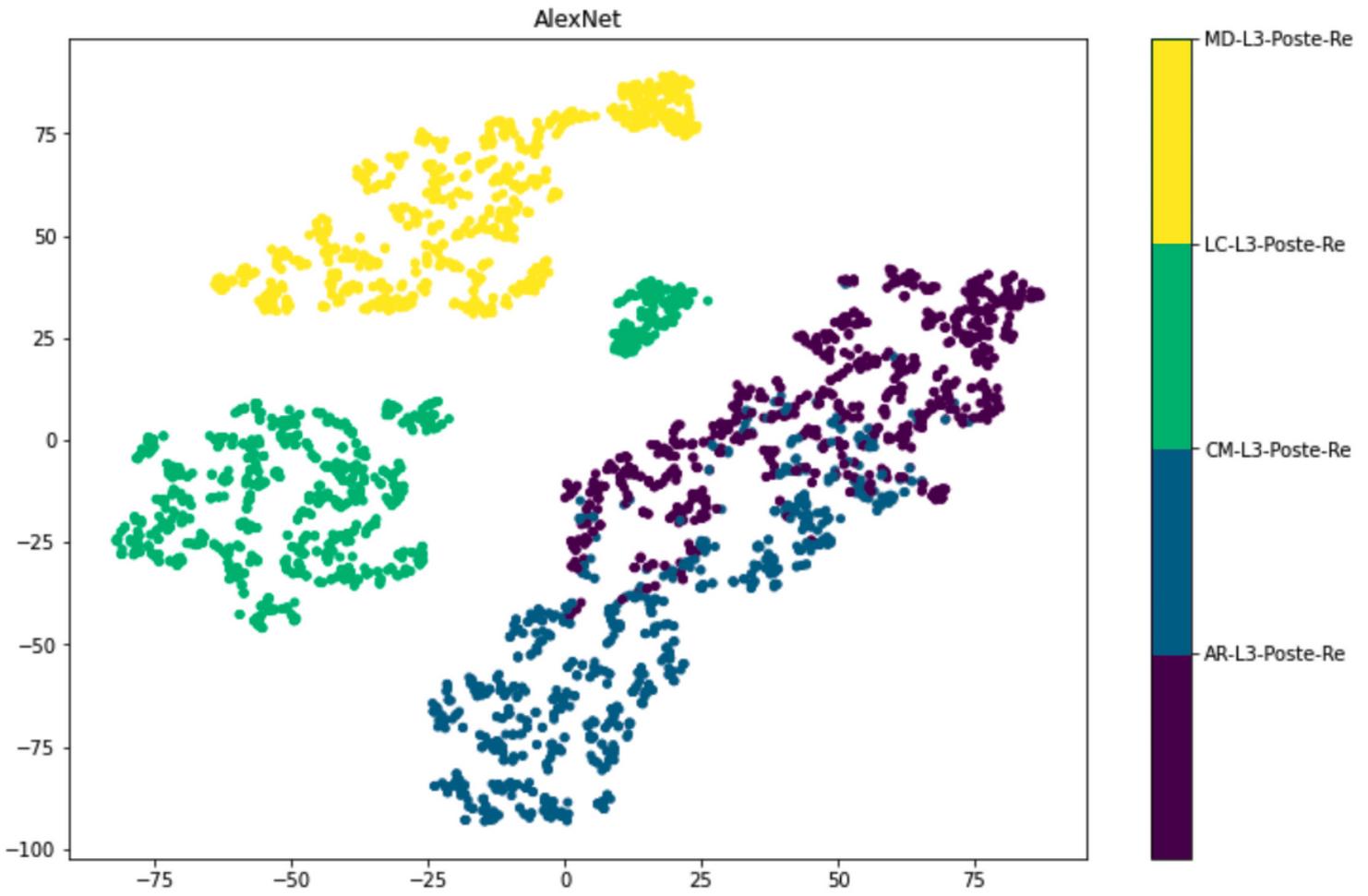


Figure 4

tSNE visualization of the AlexNet model by dimensionality reduction of the penultimate features (The test data are shown in colors for different classes.)

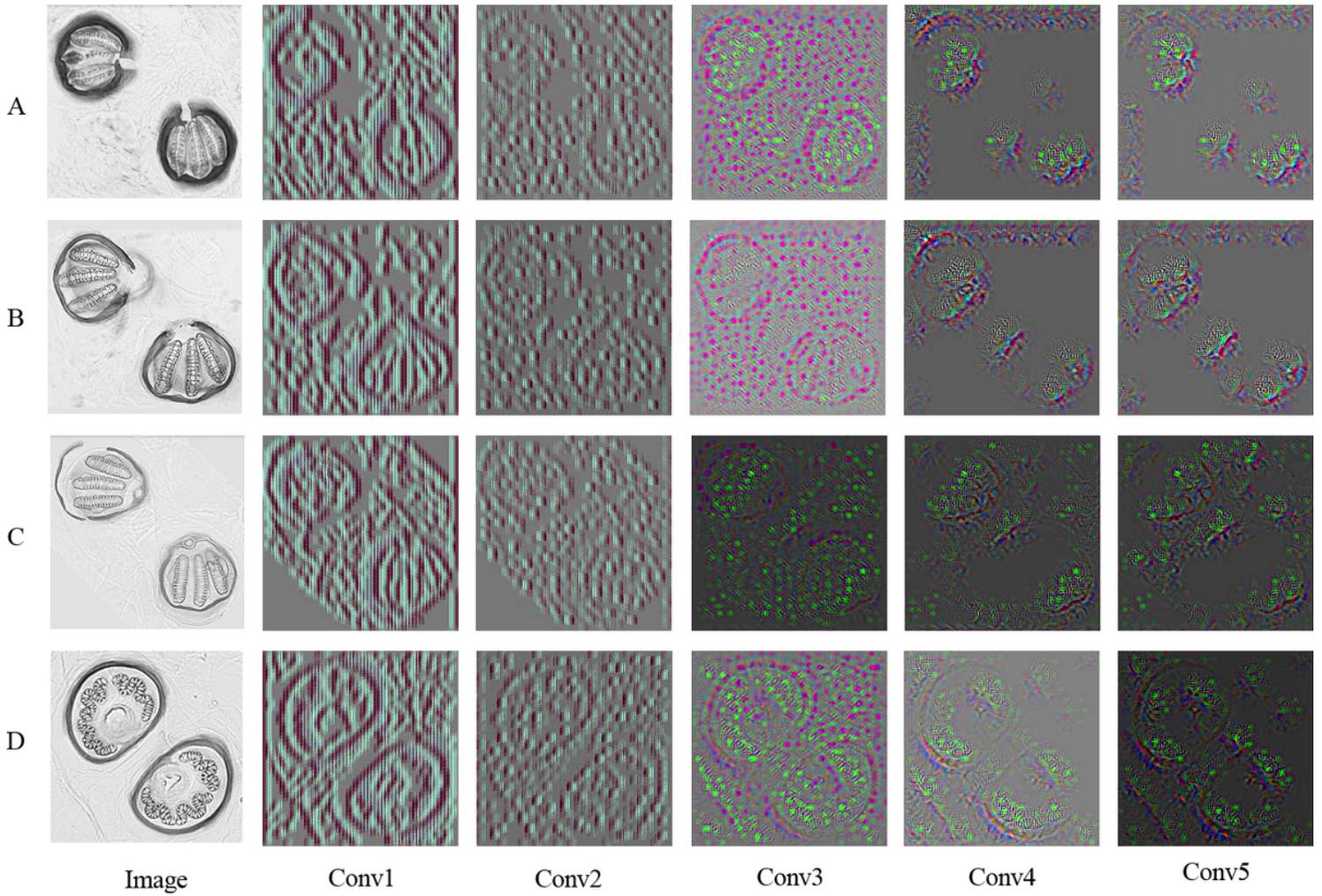


Figure 5

Visualization of hidden convolutional layers in AlexNet for four example images (To clearly show the patterns, we generated the color images; A: *Chrysomya (Achoetandrus) ruffifacies*, B: *Chrysomya megacephala*, C: *Lucilia cuprina*, D: *Musca domestica*)

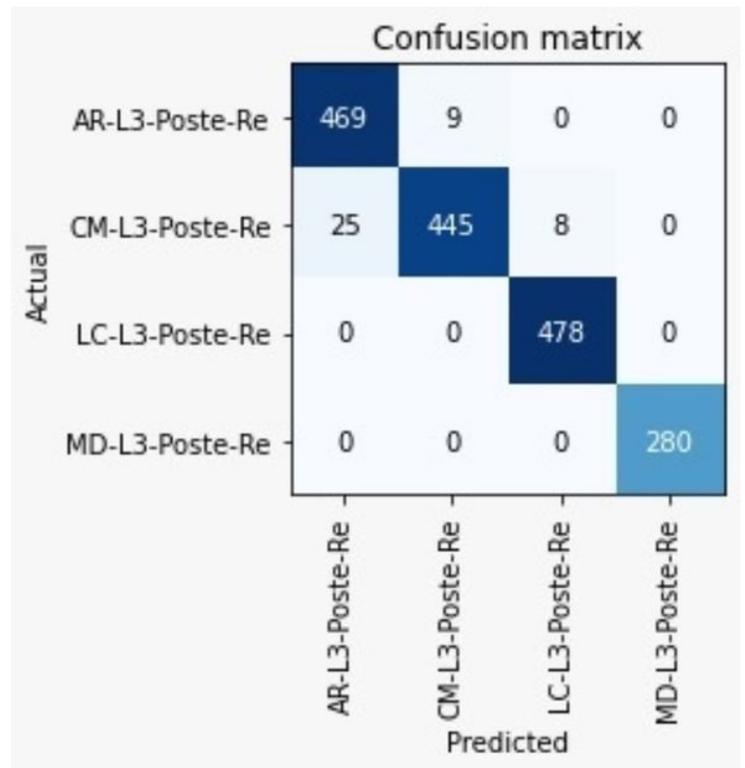
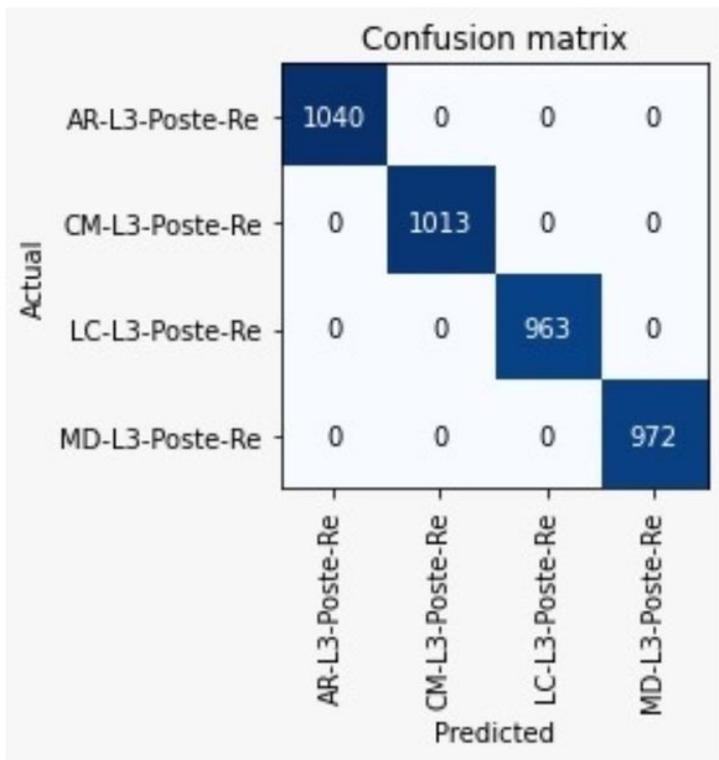
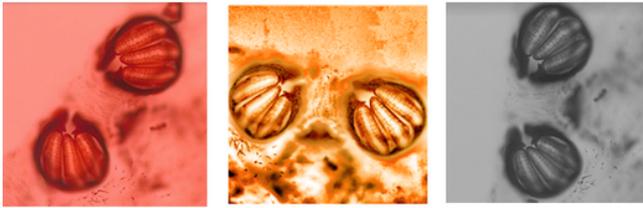


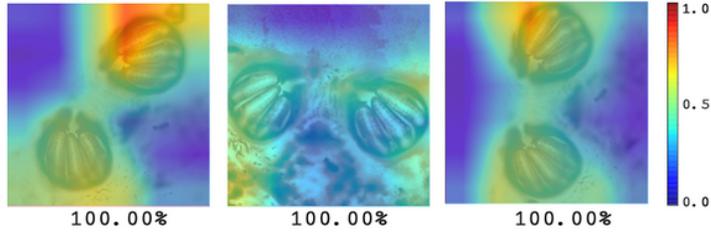
Figure 6

The confusion matrix achieved by the AlexNet model (A) Validation (B) Test dataset.

Input : *Achoetandrus ruffifacies*



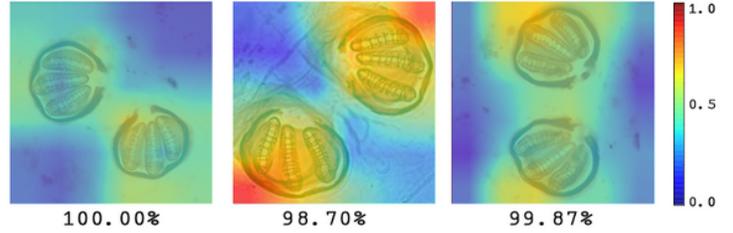
Pred : *Achoetandrus ruffifacies*



Input : *Chrysomya megacephala*



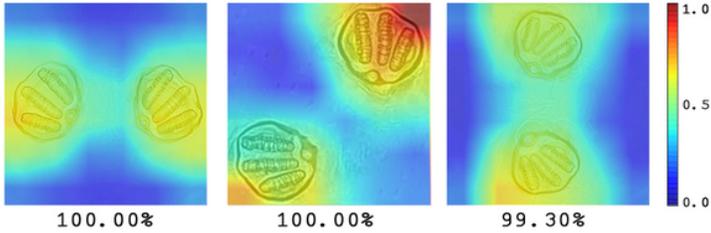
Pred : *Chrysomya megacephala*



Input : *Lucilia cuprina*



Pred : *Lucilia cuprina*



Input : *Musca domestica*



Pred : *Musca domestica*

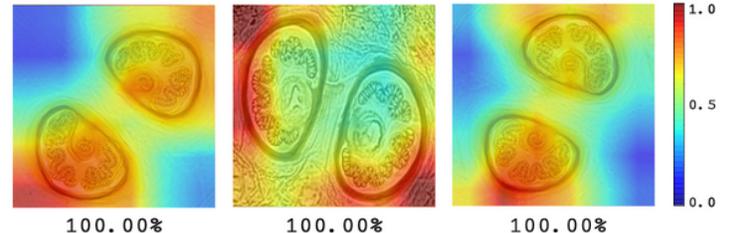


Figure 7

Heatmap of attention maps of Alexnet on example images showing prediction accuracy (98.70-100%) of this model for classification of each fly species in different image conditions

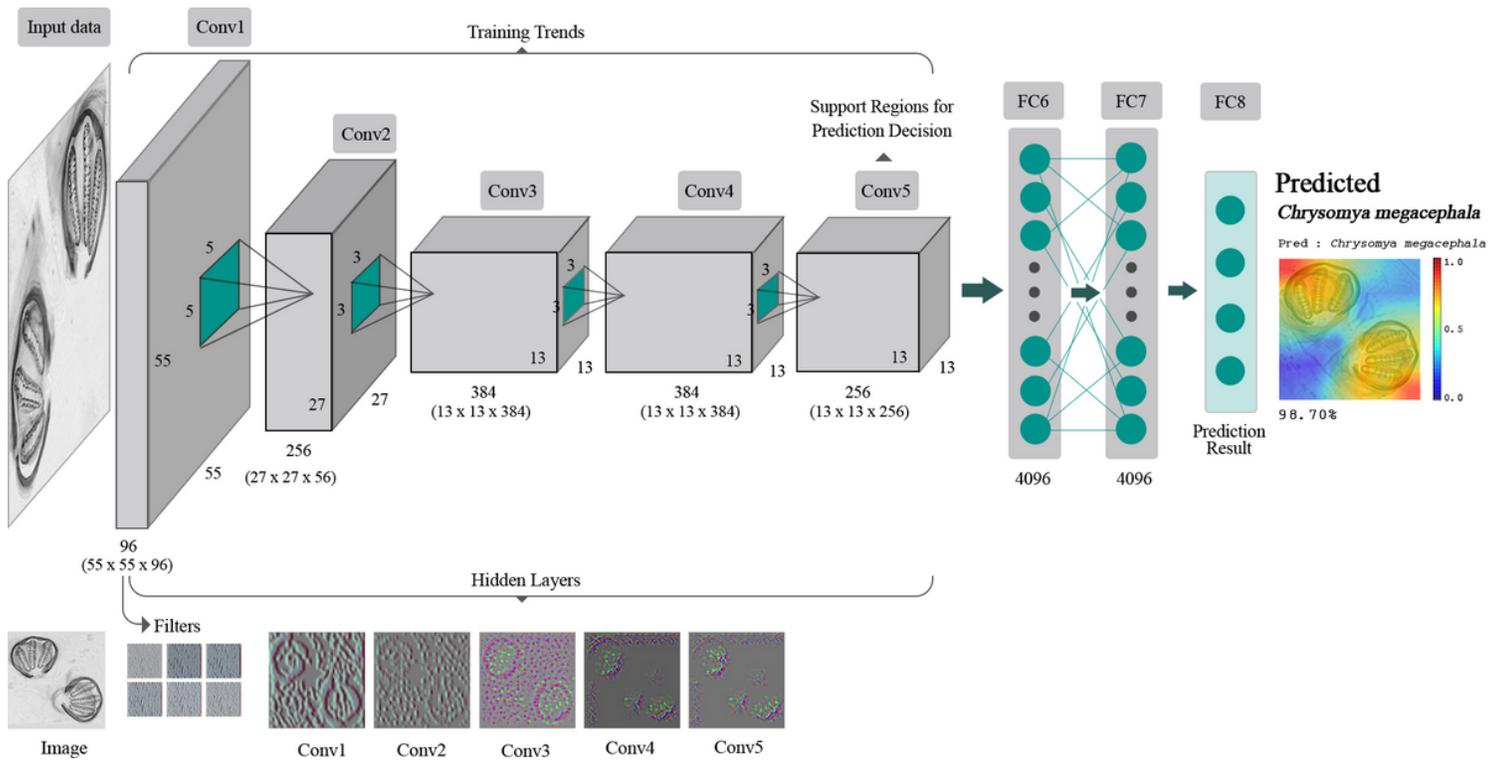


Figure 8

Framework of the proposed interpretation architecture for deep learning models, Alexnet in this study