King Saud University
College of Computer and Information Sciences
Information Technology Department

كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

IT 426 Artificial Intelligence Systems

Advanced Computer Vision with TensorFlow

Prepared by

Student Name	Student ID
Rama Khalid Alomair	444200662
Rima Khalid Alsonbul	444200524
Walaa Saif Al-eslam	444200088

Supervised by:

Dr. Reem Alqifari

Table of Contents

Course Information

Title : Advanced Computer Vision with TensorFlow

Platform : Coursera

Duration : Approximately 4 weeks / 17 hours total

Instructor : Laurence Moroney (DeepLearning.AI), Eddy Shyu (DeepLearning.AI)

Level : Intermediate to Advanced

Detailed Content Summary

This course [1] covers advanced computer vision techniques using TensorFlow, progressing from basic concepts to implementing state-of-the-art architectures for object detection and segmentation, with emphasis on model interpretation and improvement.

Key Concepts

The course covers computer vision tasks including image classification, object detection, semantic segmentation, and instance segmentation, utilizing core architectures such as the R-CNN [4] [6] family, ResNet-50 [2], U-Net [3], and Mask R-CNN.

Methods

Object detection uses RPN, Selective Search, bounding box regression with Non-Maximum Suppression, and multi-scale anchor boxes. Image segmentation employs encoder-decoder architectures [3][7] with skip connections and transposed convolution. Model interpretation includes gradient-based saliency, feature map visualization, and attention mechanisms.

AI Techniques

Core techniques include CNNs for spatial feature extraction, transfer learning from pre-trained ImageNet models [2], data augmentation for limited datasets, multi-task learning combining classification and localization, and end-to-end training with custom loss functions using gradient-based optimization.

Important Definitions

Important definitions include **bounding boxes** (rectangle coordinates defining object location), **IoU** (overlap metric between predicted and ground truth boxes), **anchor boxes** (pre-defined reference boxes with various sizes and aspect ratios), **skip connections** (direct pathways preserving information between layers), **transposed convolution** (learnable upsampling operation), **receptive field** (input region affecting neuron activation), **NMS threshold** (IoU value for suppressing overlapping detections), and **feature pyramid** (multi-scale representation for detecting objects at different sizes).

Algorithms or Frameworks

The course utilizes TensorFlow 2.x [9] with Keras API, TensorFlow Hub, and Object Detection API as primary frameworks, supported by NumPy, Matplotlib, PIL/Pillow, and OpenCV libraries. Model architectures covered include ResNet-50 [2] (deep residual network with skip connections), U-Net (encoder-decoder for segmentation) [3], Mask R-CNN (two-stage detector with segmentation) [4], SSD (fast single-stage detection), and foundational architectures VGG and AlexNet.

Applications of AI

Real-world applications

Real-world applications span multiple domains: autonomous vehicles use R-CNN, Faster R-CNN [4], and YOLO [5] for detecting pedestrians, vehicles, and obstacles in real-time navigation; healthcare systems employ object detection for identifying tumors and abnormalities in medical scans, while U-Net [3] performs organ segmentation for surgical planning and tumor boundary detection for radiation therapy; self-driving cars utilize Mask R-CNN [4][6] for road scene segmentation and lane marking identification; and explainable AI techniques like Grad-CAM [8] and saliency maps support medical diagnosis by highlighting influential image features, while researchers use visualization methods to debug models and improve architectures as demonstrated in the AlexNet case study.

Case studies or examples mentioned in the course

Course projects progress through four weeks: Week 2 features rubber ducky detection using only 5 training examples to demonstrate transfer learning with ResNet-50; Week 3 includes pet segmentation with U-Net or Mask R-CNN for pixel-level boundary detection, digit segmentation for document digitization and license plate recognition, and zombie detection teaching instance segmentation applicable to crowd analysis; Week 4 presents Fashion MNIST with Class Activation Maps to visualize features influencing predictions and AlexNet architecture analysis using layer activation visualization to understand how interpretability advances model design.

Critical Reflection

Strengths

- Practical implementation with TensorFlow Object Detection API and real-world projects.
- Excellent transfer learning coverage for achieving results with limited data.
- Diverse applications across healthcare, autonomous vehicles, and security domains.
- Strong model interpretability focus with Grad-CAM and visualization techniques.

Weaknesses

- Steep learning curve assuming substantial prior TensorFlow knowledge.
- High computational requirements with lengthy training times.
- Limited production deployment and model optimization coverage.
- Focus on older architectures without addressing recent developments.

What was missing or could be improved

- Modern architectures including Vision Transformers and attention-based models.
- Ethical considerations, bias detection, and privacy preservation.
- Model optimization techniques for deployment (quantization, pruning).
- Domain adaptation for handling real-world distribution shift.

Connections to what we've learned in this class

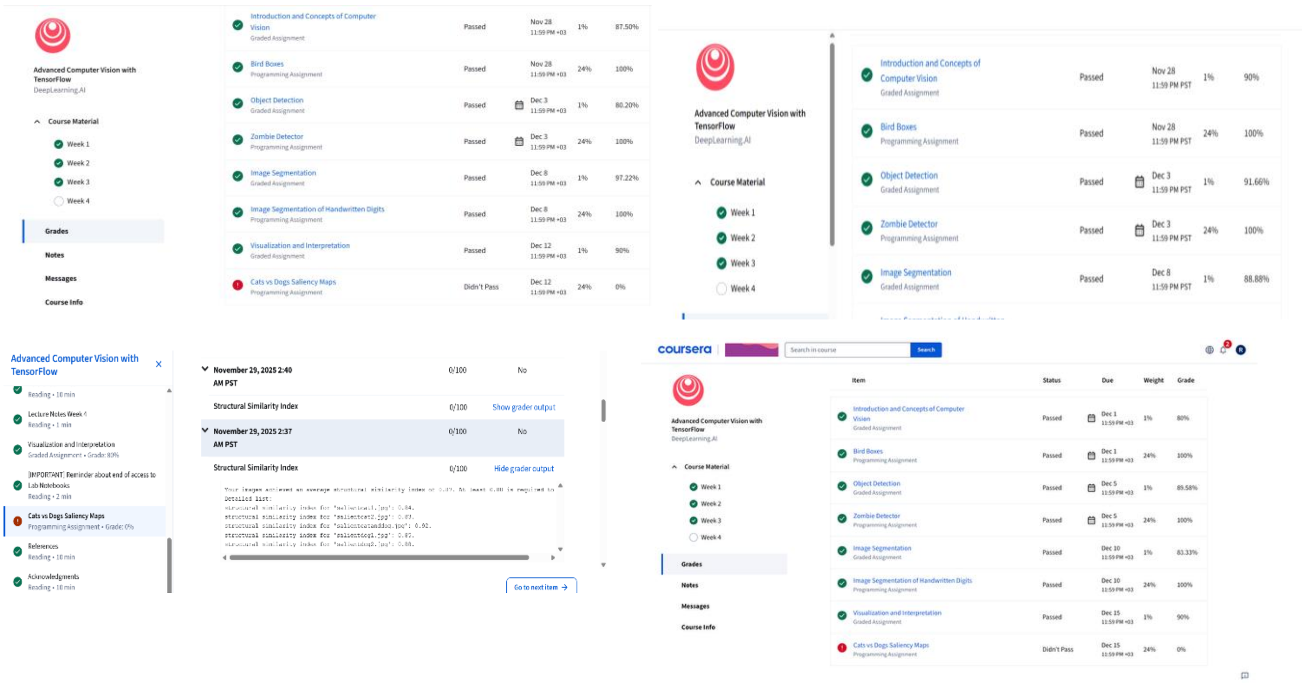
This course builds upon IT426's foundational AI concepts. IT426 introduced intelligent agents and problem-solving, which this course applied to visual understanding. The search algorithms from IT426 relate to neural network optimization and hyperparameter tuning. Knowledge representation concepts connect to hierarchical feature learning in CNNs. Both courses emphasize leveraging existing knowledge—IT426 through recommender systems, this course through transfer learning.

Personal & Group Takeaways

	perspective on what learned	Insights on how this could shape future studies or career
Rama Khalid Alomair	The hands-on experience with object detection and transfer learning provided practical skills. Model interpretability emphasized understanding predictions, not just accuracy.	I plan to pursue computer vision work in healthcare and medical image analysis.
Rima Khalid Alsonbul	Learning to decompose complex visual tasks was most valuable. Transfer learning demonstrated practical solutions with minimal data. The functional API expanded my capabilities.	I aim to specialize in computer vision for autonomous systems and robotics.
Walaa Saif Al-eslam	The course revealed the engineering complexity behind computer vision systems. Understanding detection mechanisms and	I am interested in entrepreneurial opportunities applying computer vision to agriculture and manufacturing.

	interpretability techniques was particularly valuable.	

Course Completion



We have successfully completed all the course requirements with excellence, except for the final week's assignment. The task requires achieving an accuracy above 88% for image classification, but the highest accuracy we were able to reach—after more than 30 attempts—was 87%. We have reported to Coursera that the assignment is not solvable.

References

[1] Coursera and DeepLearning.AI, "Advanced Computer Vision with TensorFlow," 2024. [Online]. Available: https://www.coursera.org/learn/advanced-computer-vision-with-tensorflow. [Accessed: Nov. 27, 2024].

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 2015, pp. 234-241.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2961-2969.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014, pp. 580-587.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626.

[9] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in Proc. 12th USENIX Symp. Operating Systems Design and Implementation (OSDI), Savannah, GA, USA, 2016, pp. 265-283.