Training Course on Transformer Models for Computer Vision

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

The landscape of **Computer Vision** has been profoundly reshaped by the advent of **Transformer Models**, particularly **Vision Transformers (ViT)**. Traditionally dominated by Convolutional Neural Networks (CNNs), the field is now experiencing a paradigm shift as ViTs demonstrate unparalleled capabilities in capturing **global context** and **long-range dependencies** within visual data. This intensive training course delves deep into the theoretical foundations and practical applications of ViT, equipping participants with the cutting-edge skills to design, implement, and deploy advanced **AI-powered solutions** for diverse **visual intelligence** tasks.

Training Course on Transformer Models for Computer Vision is meticulously designed to bridge the gap between theoretical understanding and real-world implementation, offering a comprehensive exploration of **ViT architectures**, **training methodologies**, and **optimization techniques**. Participants will gain hands-on experience with popular **deep learning frameworks** like **PyTorch** and **TensorFlow**, mastering the art of **transfer learning** and fine-tuning pre-trained ViT models for superior performance across various benchmarks. Upon completion, attendees will be proficient in leveraging ViTs for state-of-the-art **image classification**, **object detection**, **semantic segmentation**, and **generative AI** applications, positioning them at the forefront of the **AI innovation** wave.

Programme Curriculum

Training Course on Transformer Models for Computer Vision

Introduction

The landscape of **Computer Vision** has been profoundly reshaped by the advent of **Transformer Models**, particularly **Vision Transformers (ViT)**. Traditionally dominated by Convolutional Neural Networks (CNNs), the field is now experiencing a paradigm shift as ViTs demonstrate unparalleled capabilities in capturing **global**

context and **long-range dependencies** within visual data. This intensive training course delves deep into the theoretical foundations and practical applications of ViT, equipping participants with the cutting-edge skills to design, implement, and deploy advanced **AI-powered solutions** for diverse **visual intelligence** tasks.

Training Course on Transformer Models for Computer Vision is meticulously designed to bridge the gap between theoretical understanding and real-world implementation, offering a comprehensive exploration of **ViT architectures**, **training methodologies**, and **optimization techniques**. Participants will gain hands-on experience with popular **deep learning frameworks** like **PyTorch** and **TensorFlow**, mastering the art of **transfer learning** and fine-tuning pre-trained ViT models for superior performance across various benchmarks. Upon completion, attendees will be proficient in leveraging ViTs for state-of-the-art **image classification**, **object detection**, **semantic segmentation**, and **generative AI** applications, positioning them at the forefront of the **AI innovation** wave.

Course Duration

10 days

Course Objectives

1. Master the foundational concepts of Transformer architecture and self-attention mechanisms.
2. Understand the evolution from CNNs to Vision Transformers (ViT) in Computer Vision.
3. Implement ViT models from scratch using leading deep learning frameworks (PyTorch, TensorFlow).
4. Apply transfer learning and fine-tuning strategies for pre-trained ViT models.
5. Develop expertise in image classification with Vision Transformers on large-scale datasets.
6. Explore advanced object detection techniques leveraging ViTs (e.g., DETR, ViT-YOLO).
7. Gain proficiency in semantic segmentation and instance segmentation using ViT-based models.
8. Understand the role of ViTs in generative AI for image synthesis and manipulation.
9. Optimize ViT models for performance, efficiency, and deployment on various platforms.
10. Analyze and interpret ViT model behavior using explainable AI (XAI) techniques.
11. Explore hybrid architectures combining ViTs with CNNs for enhanced performance.
12. Address ethical considerations and bias in AI models for computer vision.
13. Stay abreast of trending research and future directions in Vision Transformers and large visual models.

Organizational Benefits

- Foster a workforce capable of implementing cutting-edge **AI solutions** in **computer vision**.
- Stay ahead in industries reliant on **visual data analysis** by adopting the latest **deep learning advancements**.
- Develop highly accurate and robust **computer vision systems** for automation and optimization.
- Leverage **transfer learning** and pre-trained models to accelerate project timelines.
- Equip employees with highly sought-after **AI skills**, boosting morale and career progression.
- Extract deeper insights from visual data, leading to more informed strategic choices.
- Build models capable of handling **large datasets** and complex visual tasks.

Target Audience

1. AI/ML Engineers
2. Data Scientists
3. Computer Vision Developers
4. Deep Learning Researchers
5. Image Processing Specialists

6. Software Engineers interested in AI
7. Ph.D. Candidates and Academics in related fields
8. Professionals looking to apply AI to real-world visual challenges

Course Outline

Module 1: Introduction to Transformers and Computer Vision

- Overview of Neural Networks and Deep Learning fundamentals.
- Recap of Convolutional Neural Networks (CNNs) and their limitations.
- Introduction to the original Transformer architecture from NLP.
- Motivation for applying Transformers to Computer Vision.
- Key differences and similarities between CNNs and Vision Transformers.
- **Case Study:** The rise of transformers in NLP (e.g., BERT, GPT) and the initial shift towards vision.

Module 2: Vision Transformer (ViT) Architecture

- Detailed breakdown of the ViT model structure.
- Understanding Image Patching and Linear Embeddings.
- Role of Positional Encoding in ViT.
- Exploration of the Multi-Head Self-Attention (MSA) mechanism.
- The Feed-Forward Network (MLP) and Layer Normalization.
- **Case Study:** Analysis of the "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" paper.

Module 3: Training Vision Transformers

- Strategies for effective **ViT training** from scratch.
- Importance of large-scale datasets (e.g., ImageNet) for ViT pre-training.
- Data Augmentation techniques specific to ViTs.
- Optimization algorithms and learning rate schedules for ViT.
- Regularization techniques (e.g., Dropout, Stochastic Depth) in ViT training.
- **Case Study:** Training ViT on ImageNet and benchmarking performance against CNNs.

Module 4: Transfer Learning with ViT

- Concepts of Transfer Learning and Fine-tuning in Computer Vision.
- Leveraging pre-trained ViT models for downstream tasks.
- Strategies for adapting pre-trained weights to new datasets.
- Efficient fine-tuning techniques (e.g., PEFT, LoRA for Transformers).
- Evaluating the benefits of transfer learning for **data efficiency**.
- **Case Study:** Fine-tuning a pre-trained ViT for a specialized medical imaging dataset.

Module 5: ViT for Image Classification

- Implementing ViT for image classification tasks.
- Hands-on training and evaluation on standard datasets (e.g., CIFAR-10, ImageNet subsets).
- Understanding classification head design for ViT.
- Performance metrics for image classification.
- Visualizing attention maps for interpretability.
- **Case Study:** Classifying plant diseases from images using a fine-tuned ViT.

Module 6: Advanced ViT Architectures and Variants

- Exploring hybrid ViT-CNN models.
- Swin Transformers and their shifted window attention mechanism.
- DeiT (Data-efficient image Transformers) for smaller datasets.
- MAE (Masked Autoencoders) for self-supervised learning.
- Other notable ViT derivatives and their applications.
- **Case Study:** Applying Swin Transformer for improved performance in remote sensing image classification.

Module 7: Object Detection with Vision Transformers

- Introduction to object detection challenges and traditional approaches.
- DETR (DEtection TRansformer): End-to-end object detection with Transformers.
- Understanding query learning and bipartite matching in DETR.
- Implementing ViT-based object detectors.
- Performance evaluation metrics (mAP) for object detection.
- **Case Study:** Real-time object detection in autonomous driving scenarios using DETR.

Module 8: Semantic and Instance Segmentation with ViT

- Concepts of semantic segmentation and instance segmentation.
- Adapting ViT for pixel-level prediction tasks.
- Masked Autoencoders (MAE) for segmentation pre-training.
- Panoptic Segmentation with ViT-based models.
- Applications in medical imaging and scene understanding.
- **Case Study:** Segmenting tumors in medical MRI scans using a ViT-UNet architecture.

Module 9: Generative AI and Vision Transformers

- Introduction to Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).
- Role of Transformers in image generation.
- DALL-E and other text-to-image synthesis models leveraging ViTs.
- Image-to-image translation and style transfer with ViTs.
- Future of generative models with large visual transformers.
- **Case Study:** Generating high-resolution images from textual descriptions using a ViT-powered generative model.

Module 10: Practical Implementation and Frameworks

- Hands-on coding with PyTorch and TensorFlow for ViT models.
- Leveraging Hugging Face Transformers library for quick prototyping.
- Setting up development environments (e.g., Google Colab, cloud GPUs).
- Debugging and troubleshooting common ViT implementation issues.
- Model serialization, loading, and inference.
- **Case Study:** Building a custom ViT classification pipeline using the Hugging Face library.

Module 11: Model Optimization and Deployment

- Strategies for optimizing ViT model inference speed.
- Model Quantization and Pruning for reduced memory footprint.
- Deployment of ViT models on edge devices and mobile platforms.

- Using ONNX for cross-platform model deployment.
- Monitoring and maintaining deployed ViT models.
- **Case Study:** Deploying a compact ViT model for real-time inference on a drone camera feed.

Module 12: Interpretability and Explainable AI (XAI) for ViT

- Understanding the importance of model interpretability.
- Techniques for visualizing attention maps in ViT.
- Saliency maps and Grad-CAM for understanding ViT predictions.
- Analyzing patch importance and feature representation.
- Addressing bias and fairness in ViT models.
- **Case Study:** Explaining a ViT's decision-making process for identifying defects in manufacturing.

Module 13: ViT in Real-World Applications (Advanced)

- Autonomous Driving: Perception, scene understanding, and prediction with ViT.
- Medical Imaging: Diagnostics, disease detection, and surgical assistance.
- Remote Sensing: Land cover classification, disaster monitoring, and urban planning.
- Retail and E-commerce: Visual search, product recommendation, and quality control.
- Security and Surveillance: Anomaly detection and facial recognition.
- **Case Study:** Developing a ViT-based system for traffic sign recognition in autonomous vehicles.

Module 14: Challenges and Limitations of ViT

- Data Efficiency of ViT compared to CNNs.
- Computational cost and memory footprint of large ViTs.
- Inductive biases and their impact on ViT performance.
- Current research directions to mitigate ViT limitations.
- The interplay between model size, data, and performance.
- **Case Study:** Analyzing the challenges of training large ViTs on limited computational resources.

Module 15: Future Trends and Research Directions

- Foundation Models in Computer Vision beyond ViT.
- Multimodal Transformers (e.g., combining vision and language).
- Efficient Attention Mechanisms and their impact.
- The role of self-supervised learning in future ViT development.
- Emerging research frontiers in large visual models.
- **Case Study:** Discussion on the potential of multimodal transformers for advanced human-computer interaction.

Training Methodology

Our training methodology emphasizes a **blended learning approach**, combining theoretical instruction with extensive hands-on practice.

- **Interactive Lectures:** Engaging presentations explaining complex concepts with real-world examples.
- **Live Coding Sessions:** Step-by-step demonstrations of ViT implementation in PyTorch and TensorFlow.
- **Hands-on Labs & Exercises:** Practical assignments to solidify understanding and build proficiency.
- **Case Study Analysis:** In-depth examination of real-world applications and their challenges.
- **Project-Based Learning:** Participants will work on a capstone project to apply learned skills.

- **Group Discussions & Peer Learning:** Collaborative problem-solving and knowledge sharing.
- **Q&A Sessions:** Dedicated time for addressing participant queries and clarifying doubts.
- **Mentorship & Support:** Access to experienced instructors for personalized guidance.
- **Practical Tooling:** Emphasis on using industry-standard libraries and frameworks (Hugging Face, PyTorch Lightning).

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com