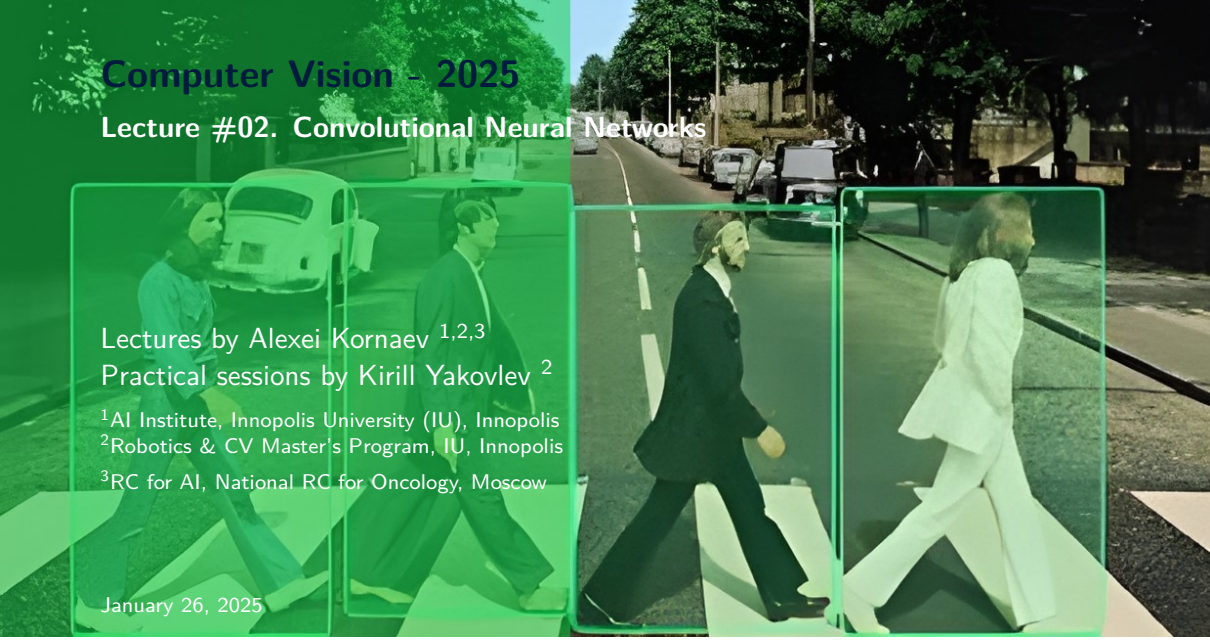
Computer Vision - 2025

Lecture #02. Convolutional Neural Networks



Lectures by Alexei Kornaev [1,2,3]

Practical sessions by Kirill Yakovlev [2]

[1]AI Institute, Innopolis University (IU), Innopolis

[2]Robotics & CV Master's Program, IU, Innopolis

[3]RC for AI, National RC for Oncology, Moscow

January 26, 2025

Agenda

1. Learning Outcomes and Content of the Course
2. Place and Role of CV in Our Days
3. Deterministic Models in CV
4. Support of Coding
5. Conclusion

Section 1. Learning Outcomes and Content of the Course

Outcomes

The course is dedicated to studying key ideas and modern methods of CV. It covers topics ranging from basic image processing techniques to advanced deep learning architectures, including convolutional neural networks (CNNs), vision transformers (ViTs), generative models, and multimodal data processing. The course objectives for the students are:

1. Learn how CV models work: from basic ideas to real-world problems
2. Learn how to build a model from scratch using LLM assistances or using an open-source model to solve a problem
3. Take a few more steps forward from educational problems to scientific ones

Outcomes

Level 1: What concepts should the student understand / know / remember / explain by the end of the course?

1. Basic principles of image formation, filtering, and feature extraction
2. Intuition behind architectures and training of CV models
3. Principles of generative models
4. Fundamentals of object detection, recognition, segmentation, and depth estimation techniques
5. Challenges and methods for processing 3D images, point clouds, and video data

Outcomes

Level 2: What practical skills should the student possess by the end of the course?

1. Preprocess and augment image data
2. Implement and train CNNs for image classification and object detection
3. Fine-tune pre-trained models for custom tasks using transfer learning
4. Train models to generate realistic images or perform image reconstruction
5. Implement segmentation and object detection models
6. Perform depth estimation and landmark detection
7. Build face recognition systems
8. Process 3D images, point clouds, and video data

Outcomes

Level 3: What complex skills should the student be able to apply in real-life situations?

1. Conduct a review of modern CV methods and tools, and identify the best solutions for given tasks
2. Implement CV tasks as software code using modern libraries
3. Apply transfer learning, ensemble methods, and fine-tuning to achieve the best metrics
4. Solve complex tasks that require combining multiple types of computer vision models, such as multimodal data processing (e.g., image and text)
5. Plan and conduct computational experiments, process the results, and perform precise tuning of computer vision models
6. Deploy computer vision models in real-world applications, ensuring robustness and scalability in production environments

Content

The course content is available on both Moodle and GitHub:

1. Moodle
2. Git project [CV-2025]

Section 2. Place and Role of CV in Our Days

AI Index in Russia by MIPT (# 13, 2023)

figures/AI_index_in_Russia_2023.png

Place and Role of CV in Our Days

figures/RBWeide.png

Section 3. Deterministic Models in CV

Introduction to Images in Computer Vision

What is an Image in CV?

- An image is a numerical representation of visual information.
- Typically stored as a 2D grid of pixels, each representing intensity or color.
- Contains spatial information that can be processed by computers.

Key Elements of an Image:

- **Pixel:** Smallest unit of an image.
- **Resolution:** Dimensions of the image grid (e.g., 1920×1080 pixels).
- **Color Channels:** Defines color information (e.g., grayscale, RGB, multispectral).

Types of Images in CV

1. **Grayscale Image:**
   - Single-channel image with intensity values ranging from 0 (black) to 255 (white).
2. **Color Image:**
   - Multiple channels representing color information (e.g., Red, Green, Blue in RGB format).
3. **Binary Image:**
   - Each pixel is either 0 (black) or 1 (white), commonly used in thresholding.
4. **Multispectral/Hyperspectral Image:**
   - Captures data beyond visible light, such as infrared or X-rays.

Key Attributes of an Image

1. **Resolution:**
   - Defined as the width $\times$ height of the image (e.g., 1280×720).
   - Higher resolution provides more detail but increases computational cost.

2. **Bit Depth:**
   - Determines the range of intensity values for each pixel (e.g., 8-bit, 16-bit).

3. **Image Storage Formats:**
   - Common formats include JPEG, PNG, BMP, and TIFF.

Workflow of Image Processing in CV

Steps in Using Images for Computer Vision Tasks:

1. **Preprocessing:** Resizing, normalization, noise reduction.
2. **Feature Extraction:** Detecting edges, textures, or patterns.
3. **Analysis:** Tasks like object detection, segmentation, and classification.
4. **Advanced Applications:** Face recognition, depth estimation, and 3D processing.

Practical Tools:

- Libraries: OpenCV, PIL, scikit-image.
- Deep Learning Frameworks: TensorFlow, PyTorch.

Hands-on with the book by Howard and Gugger [2020]

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	29	150	195	254	255	254	176	193	150	96	0	0	0
2	0	0	0	48	166	224	253	253	234	196	253	253	253	253	233	0	0	0
3	0	93	244	249	253	187	46	10	8	4	10	194	253	253	233	0	0	0
4	0	107	253	253	230	48	0	0	0	0	0	192	253	253	156	0	0	0
5	0	3	20	20	15	0	0	0	0	0	43	224	253	245	74	0	0	0
6	0	0	0	0	0	0	0	0	0	0	249	253	245	126	0	0	0	0
7	0	0	0	0	0	0	0	14	101	223	253	248	124	0	0	0	0	0
8	0	0	0	0	0	11	166	239	253	253	253	187	30	0	0	0	0	0
9	0	0	0	0	0	16	248	250	253	253	253	253	232	213	111	2	0	0
10	0	0	0	0	0	0	0	43	98	98	208	253	253	253	253	187	22	0

Figure: Check the code: 04 mnist basics

Section 4. Support of Coding

Support of Coding

figures/Flowchart4Coding.jpg

Support of Coding

Two frameworks should be met:

1. Review the Lightning framework (Level Up, Core API, Optional API sections of the manual)
2. Review the ClearML documentation

Section 5. Conclusion

Conclusion

Summary:

- Images in computer vision are numerical representations of visual data, stored as pixel grids.
- Different types of images (grayscale, color, binary, multispectral) serve various purposes in CV applications.
- Key attributes like resolution, bit depth, and storage formats play critical roles in image processing.

Key Takeaway:

- Understanding image structures and properties is fundamental to developing effective CV models and applications.

Bibliography

CV-2025. The course repository. https://github.com/Mechanics-Mechatronics-and-Robotics/CV-2025, 2025. Accessed: January 26, 2025.

J. Howard and S. Gugger. **Deep Learning for Coders with Fastai and Pytorch: AI Applications Without a PhD**. O'Reilly Media, Incorporated, 2020. ISBN 9781492045526. URL https://books.google.no/books?id=xd6LxgEACAAJ.