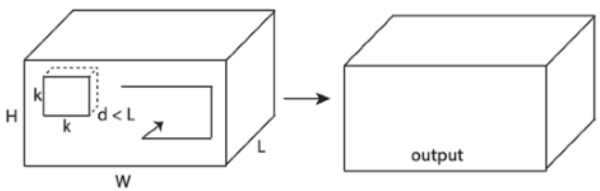
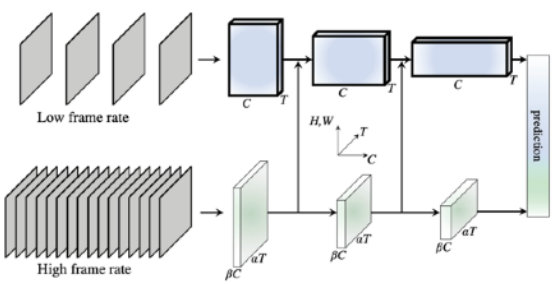
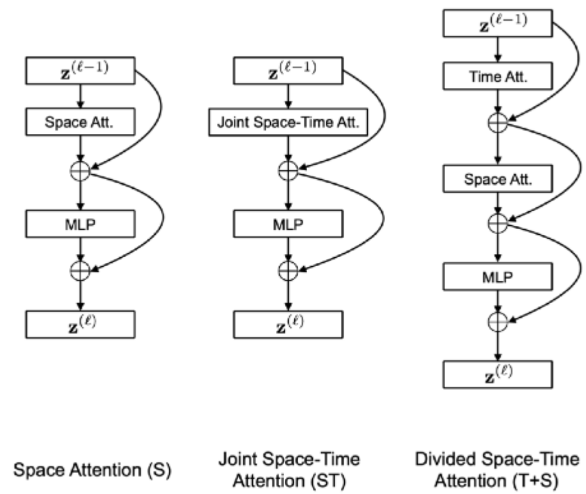
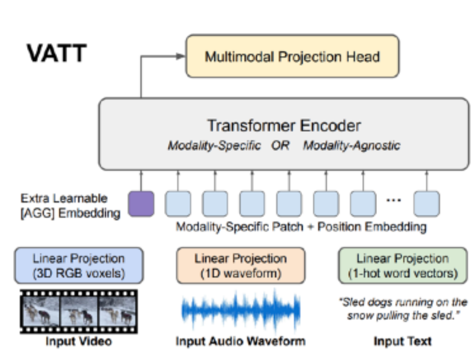
COMP 590/790 Video Recognition









https://www.gedasbertasius.com/comp790-23f

Course Introduction

Gedas Bertasius

About Me

- Originally from Lithuania.
- Came to the US to play basketball.
- Got a PhD from UPenn.
- Spent 2 years at Facebook AI Research.
- Joined UNC in 2021.

Introductions in Canvas

- Name?
- BA / BS / MS / PhD?
- Year?
- What are you excited about in computer vision and AI in general?
- Why are you taking this course?

Plan for Today

1. Motivation for the course

2. Course logistics

3. Overview of convolutional neural networks for images

Plan for Today

1. Motivation for the course

2. Course logistics

3. Overview of convolutional neural networks for images

Video Recognition

What does it mean for machines to understand video?

Action recognition

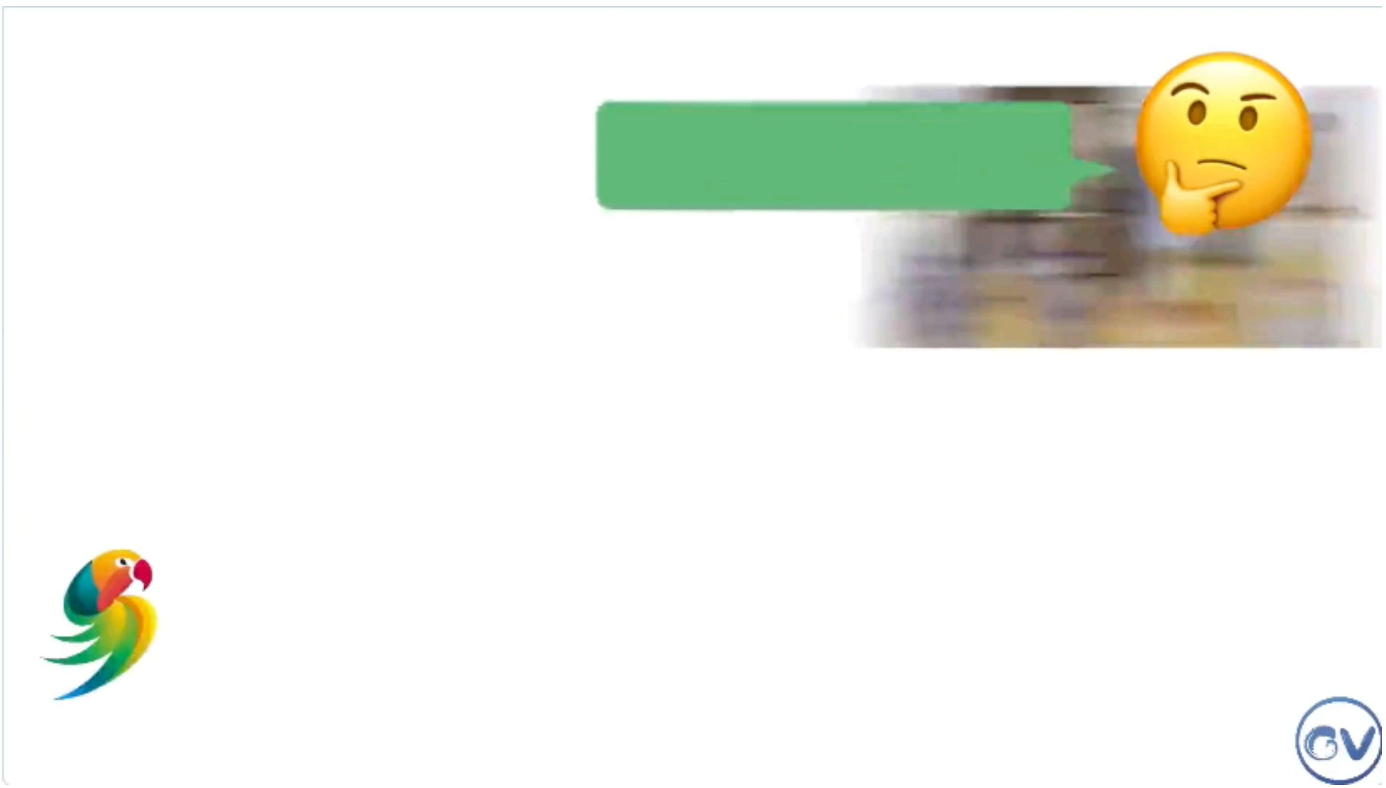
Video Recognition

What does it mean for machines to understand video?

Object detection and tracking

Video Recognition

What does it mean for machines to understand video?



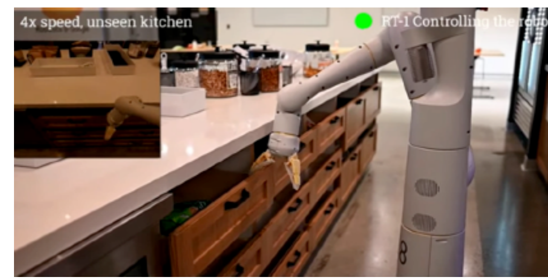
Video captioning

Why Video?

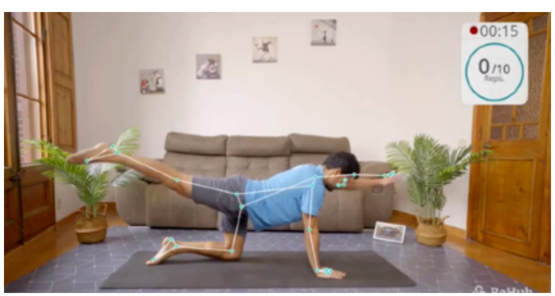
- Video recognition can have a significant impact on many real-world applications.

Augmented Reality

Self-Driving Cars

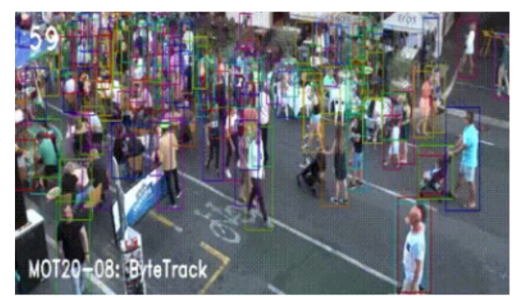


Autonomous Robots



Health

Ecology



Security

Why Video?

- **3.1 billion people** consume videos on the Internet daily.

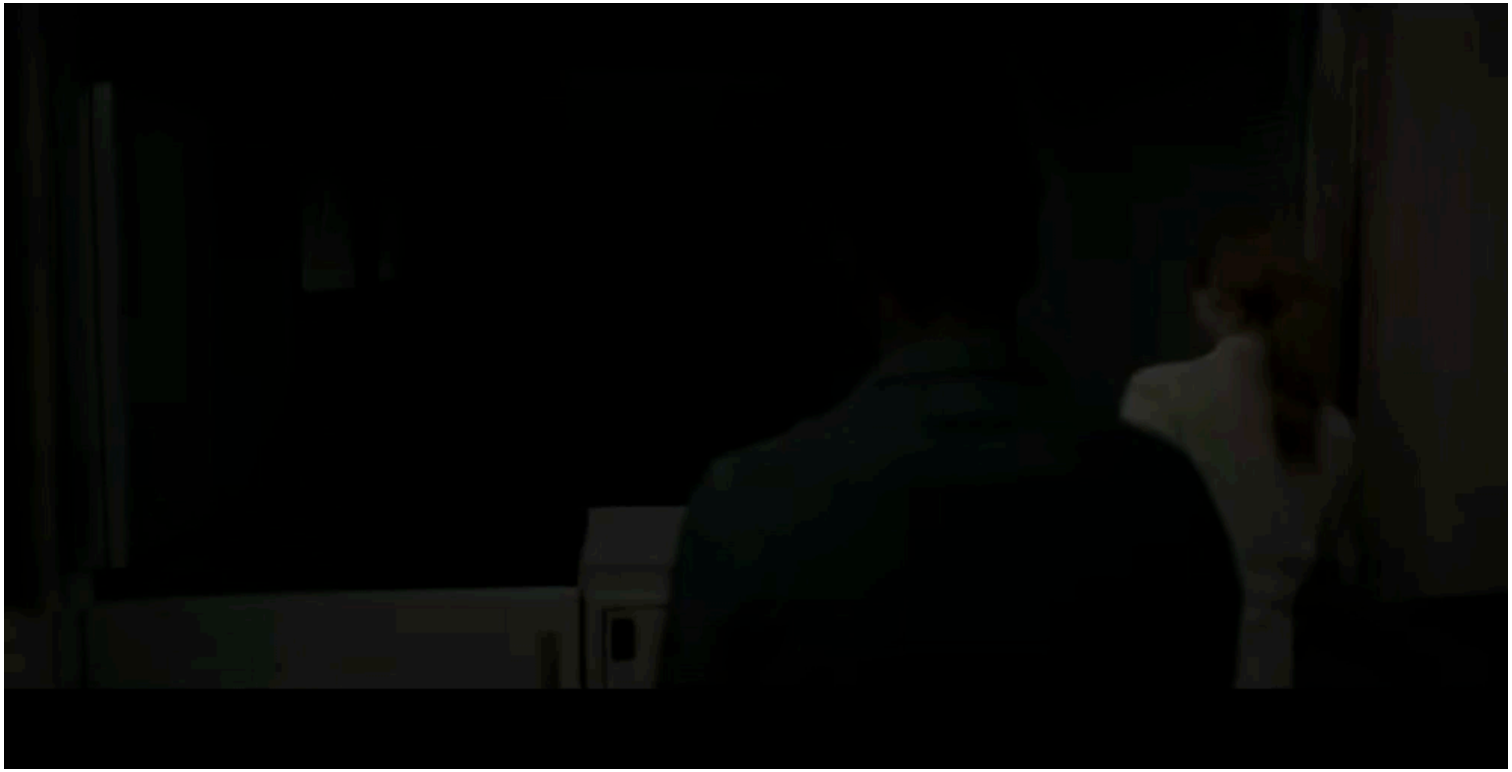
- YouTube users watch **1 billion hours (114K years)** of video daily.

- TikTok has over **1.53 billion** monthly active users.

- **720K hours (82 years)** of video are uploaded to YouTube every day.

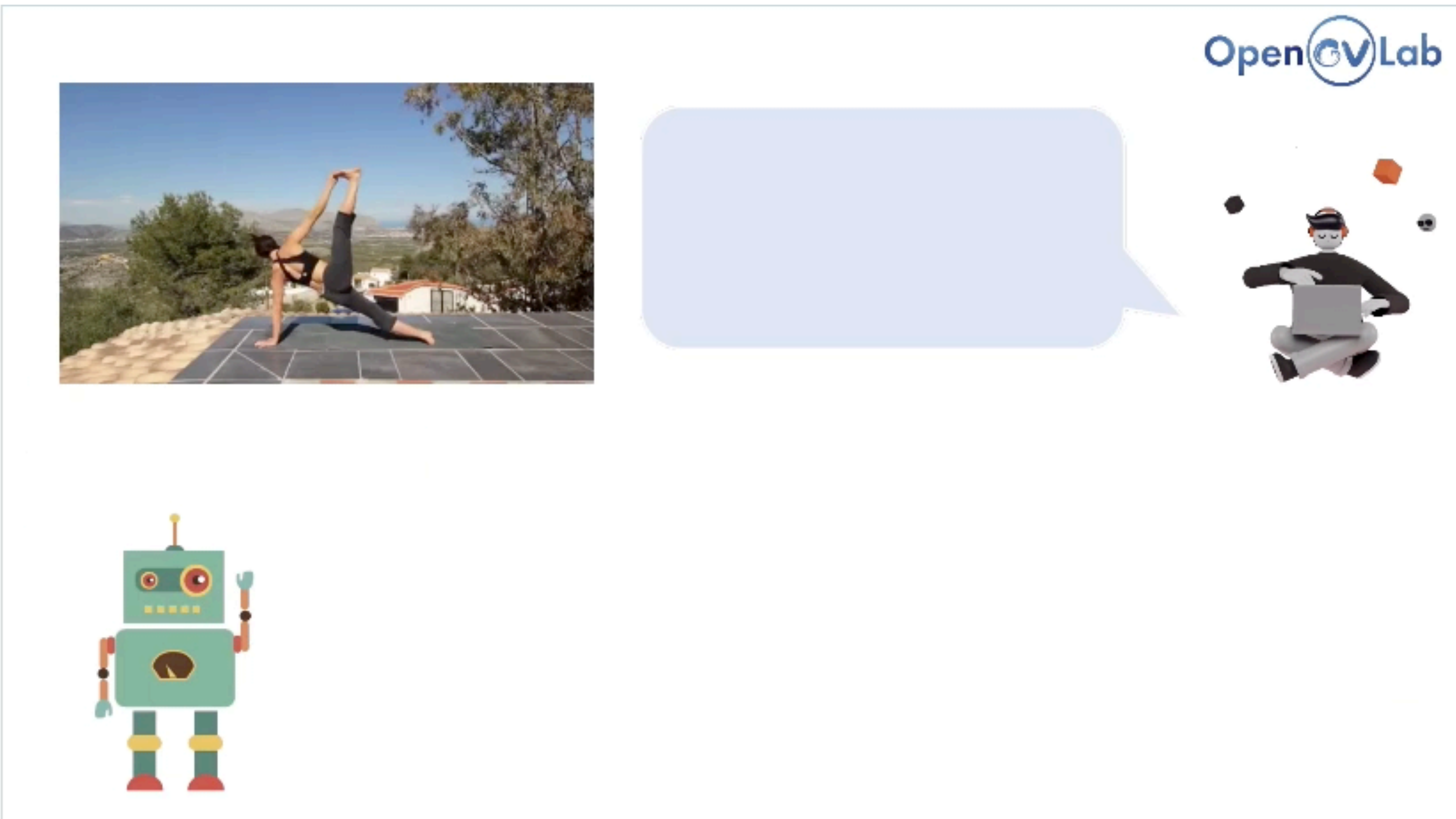
Why Video?

- Compared to images, videos contain multiple modalities (audio, speech, etc.) and complex temporal structure.



Why Video?

- Many exciting recent developments in the video domain.



Video Chatbot

Why Video?

- Many exciting recent developments in the video domain.

AI-generated Video

Why Video?

- Many exciting recent developments in the video domain.

Text-based Video Editing

Plan for Today

1. Motivation for the course

2. Course logistics

3. Overview of convolutional neural networks for images

Course Objectives

- Provide a thorough overview of state-of-the-art in this area of research.
- Learn how to analyze and present research papers.
- Discuss cutting edge research and speculate about future research directions.
- Carry out a semester-long project.

Prerequisites

- Understanding of fundamental machine learning concepts.
- Experience with deep learning / computer vision.
- The ability to analyze research papers published in major machine learning and computer vision conferences.
- If you are not sure about your background, check the papers listed on the course website, and see if you would be comfortable presenting them (also come talk to me).

Grading

- Class Participation: 10%
- Written Paper Critiques: 20%
- Paper Presentations: 30%
- Course Project: 40%

Paper Presentations

- Everybody will give two types of presentations:
  1. One 30min or one 45min paper presentation (presented solo or in pairs).
  2. One 20min paper + discussion for a paper battle (presented in groups of ~3).

- Rehearse your talk to make sure it fits within the time limit listed in the Schedule next to each paper.

- If you need help understanding the paper, send me an email and we'll set up a meeting.

Paper Presentations

- Focus on the most important high-level concepts. No need to present every single technical detail / experiment.

- Your audience should understand:
    1. The research problem.
    2. The motivation of the proposed research.
    3. Any necessary background info.
    4. The main technical details.
    5. The key experimental results.

- Spend time on description of the experiments.

Written Paper Critiques

- The goal is to provide a critical analysis of the paper (positive or negative).
- You will need to submit paper critiques for 10 of my selected papers (each critique worth 2% of the total course grade).
- Each critique graded as Pass or Fail (no detailed feedback).
- Use the template here (also provided in Canvas).
- Please write your critiques independently.
- Upload the critiques in a PDF format on Canvas by 1:00 PM on the day of the class.

QA Prompts for a Paper Discussion

- With each paper critique, you will also include one paper discussion question and your answer to that question.
- We will use your submitted discussion questions for detailed ~30min paper discussions.
- Check out some general info here on how to come up with good questions for a discussion.
- I will read every single one of these so you should come up with meaningful questions.

Detailed Paper Discussions

- For 10 of the selected papers, we will have detailed ~30min paper discussions.
- I will use your submitted QA discussion prompts (from your paper critiques) to compile a set of 6-8 discussion questions.
- We will then break out into small groups where each group will discuss one of the questions among themselves.
- Afterward, we will reconvene to discuss all of the questions together.

Paper Battles

- For 5 paper pairs, we will have paper battles, i.e., detailed head-to-head paper comparisons between two groups of students.

- Assume that the two given papers represent two conference paper submissions.

- However, only one of those papers can be accepted.

- Two groups of students will try to convince the audience that their presented paper should be “accepted”.

Paper Battles (Continued)

- Each group will give a brief 20min overview of their assigned paper (time limit will be strictly enforced).
- Following both presentations, each group will present slides with 3 main reasons why their presented paper is better.
- Afterward, we will have a brief discussion allowing each group to rebut another group's points.
- Lastly, the students in the class who were not presenting will vote on which paper is better (and provide a justification for their vote).

Course Projects

- There will be two tracks for course projects: (1) graduate and (2) undergraduate.
- Projects should be completed in groups of 2-3.
- If you want to pursue a project individually (e.g., for your dissertation, etc.), please talk to me before doing so.
- If you do not have access to GPUs, send me an email or talk to me after class.
- Start thinking early!

Graduate Track

- Aimed at graduate students pursuing research requiring video analysis.
- You can propose any project involving video recognition in any area of interest to you.
- Review the paper list for inspiration, or come talk to me for topic related suggestions.
- Undergraduate students interested in video recognition research are welcome to pursue this track but should talk to me before doing so.

Undergraduate Track

- Aimed at undergraduate students who want to gain more experience with existing video recognition tools.
- The project will require students applying existing video recognition tools to 5 of their selected video applications.
- For each application, the students will need to identify 3 failure and 3 success instances and document their experiences/insights about the model's shortcomings, etc.
- Graduate students are also welcome to pursue this track if it fits them better.

Undergraduate Track

Potentially useful pointers of existing video tools:

- Video Classification / Action Recognition
- Object Detection and Tracking
- Object Segmentation and Tracking
- Video Captioning / QA
- Audiovisual Video Captioning / QA
- Text-to-Video Generation

Project Submissions

- Proposal (10% of the total grade):
  - Presentations on **10/02/23** & **10/04/23**.
  - Write-up due **10/08/23.**
  - Overview of your project plan.

- Milestone (10% of the total grade):
  - Presentations on **10/30/23** & **11/01/23**.
  - Write-up due **11/05/23.**
  - A checkpoint to make sure you are making progress.

- Final (20% of the total grade):
  - Presentations on **12/04/23** & **12/06/23**.
  - Write-up due **12/10/23.**
  - Final findings of the project.

- Use templates here for your project write-up (also available on Canvas).
- Upload the PDF of your slides & write-ups on Canvas in the Assignment section.

Office Hours

- Office hours are available by appointment.
- Please sign up for a meeting at https://calendly.com/gedasb
- If none of the time slots work, send me an email to schedule an appointment separately.

Canvas

- We will use Canvas for many course-related activities.
- All the announcements will be made on Canvas so please check it regularly.
- You will need to upload your assignments on Canvas.
- The discussion and collaboration pages are enabled on Canvas. Please share any interesting papers, blog posts, or general ideas in the discussion page.
- You can find collaborators for project on Canvas as well.

Plan for Today

1. Motivation for the course

2. Course logistics

3. Overview of convolutional neural networks for images

Videos as Images

A video can be viewed as a collection of images.

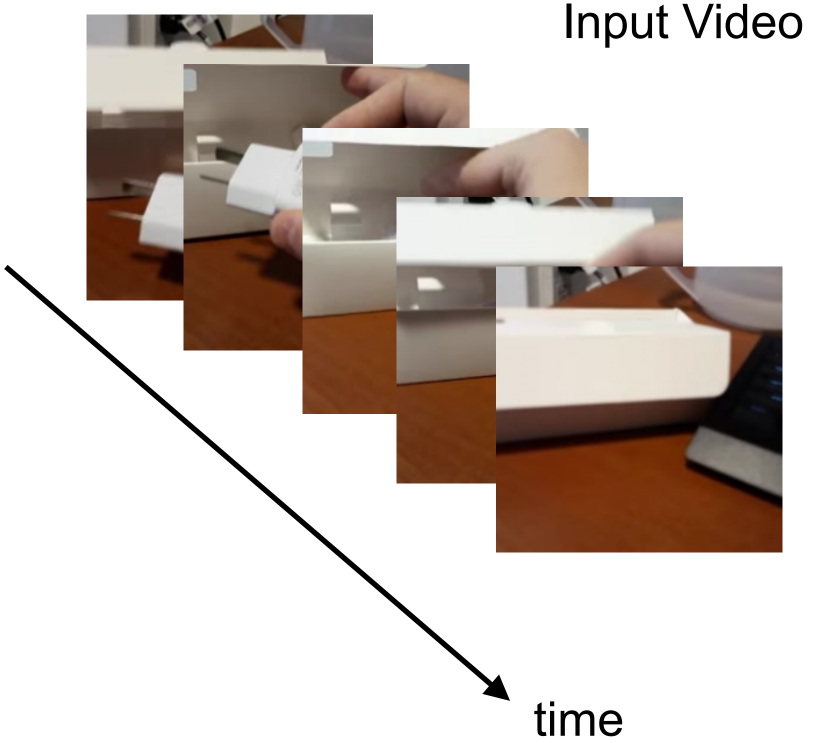
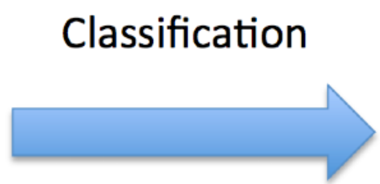


Image Classification

- The goal is to identify the category of a given image.

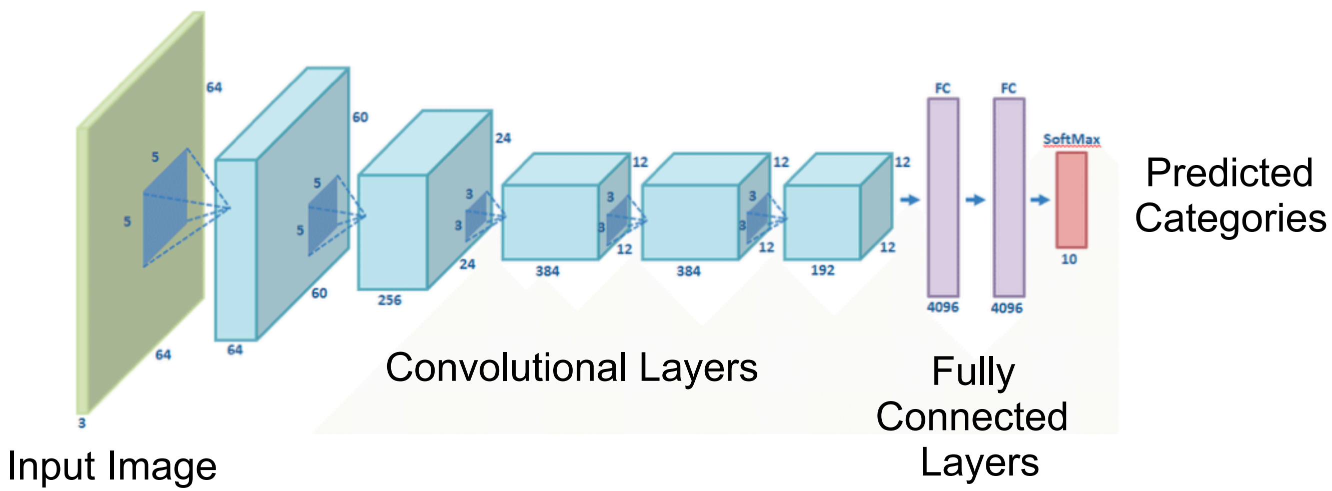
Input:



Output:

Convolutional Networks

- **Input:** An image.
- **Convolutional Layers:** a series of 2D convolutions applied one after the other.
- **Fully Connected Layers:** layers where all the units from one layer are connected to every unit of the next layer.
- **Output:** probabilities for each predicted category.



2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 0 & 5 & 6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline {}^{1}0 & {}^{2}0 & {}^{1}0 & 0 & 0 & 0 \\ \hline {}^{0}0 & {}^{0}0 & {}^{0}0 & 0 & 0 & 0 \\ \hline {}^{1}0 & {}^{2}0 & {}^{1}1 & 2 & 3 & 4 \\ \hline 0 & 0 & 5 & 6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 & 0 & 0 \\ \hline 0 & {}^{0}0 & {}^{0}0 & {}^{0}0 & 0 & 0 \\ \hline 0 & {}^{1}0 & {}^{2}1 & {}^{1}2 & 3 & 4 \\ \hline 0 & 0 & 5 & 6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 & 0 \\ \hline 0 & 0 & {}^{0}0 & {}^{0}0 & {}^{0}0 & 0 \\ \hline 0 & 0 & {}^{1}1 & {}^{2}2 & {}^{1}3 & 4 \\ \hline 0 & 0 & 5 & 6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|} \hline 0 & 0 & 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 \\ \hline 0 & 0 & 0 & {}^{0}0 & {}^{0}0 & {}^{0}0 \\ \hline 0 & 0 & 1 & {}^{1}2 & {}^{2}3 & {}^{1}4 \\ \hline 0 & 0 & 5 & 6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & 4 & 8 & 12 \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$f =$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

convolution filter

$g =$

1	2	1
0	0	0
1	2	1

$h = g * f =$

0	0	0	0	0	0
$^{1}0$	$^{2}0$	$^{1}0$	0	0	0
$^{0}0$	$^{0}0$	$^{0}1$	2	3	4
$^{1}0$	$^{2}0$	$^{1}5$	6	7	8
0	0	9	10	11	12
0	0	13	14	15	16

$=$

1	4	8	12
5			

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 & 0 & 0 \\ \hline 0 & {}^{0}0 & {}^{0}1 & {}^{0}2 & 3 & 4 \\ \hline 0 & {}^{1}0 & {}^{2}5 & {}^{1}6 & 7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 & 0 \\ \hline 0 & 0 & {}^{0}1 & {}^{0}2 & {}^{0}3 & 4 \\ \hline 0 & 0 & {}^{1}5 & {}^{2}6 & {}^{1}7 & 8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want
to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & {}^{1}0 & {}^{2}0 & {}^{1}0 \\ \hline 0 & 0 & 1 & {}^{0}2 & {}^{0}3 & {}^{0}4 \\ \hline 0 & 0 & 5 & {}^{1}6 & {}^{2}7 & {}^{1}8 \\ \hline 0 & 0 & 9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & 28 \\ \hline & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline {}^{1}0 & {}^{2}0 & {}^{1}1 & 2 & 3 & 4 \\ \hline {}^{0}0 & {}^{0}0 & {}^{0}5 & 6 & 7 & 8 \\ \hline {}^{1}0 & {}^{2}0 & {}^{1}9 & 10 & 11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & 28 \\ \hline 10 & & & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$f =$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

convolution filter

$g =$

1	2	1
0	0	0
1	2	1

$h = g * f =$

0	0	0	0	0	0
0	0	0	0	0	0
0	$^{1}0$	$^{2}1$	$^{1}2$	3	4
0	$^{0}0$	$^{0}5$	$^{0}6$	7	8
0	$^{1}0$	$^{2}9$	$^{1}10$	11	12
0	0	13	14	15	16

$=$

1	4	8	12
5	16	24	28
10	32		

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & {}^{1}1 & {}^{2}2 & {}^{1}3 & 4 \\ \hline 0 & 0 & {}^{0}5 & {}^{0}6 & {}^{0}7 & 8 \\ \hline 0 & 0 & {}^{1}9 & {}^{2}10 & {}^{1}11 & 12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & 28 \\ \hline 10 & 32 & 48 & \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & {}^{1}2 & {}^{2}3 & {}^{1}4 \\ \hline 0 & 0 & 5 & {}^{0}6 & {}^{0}7 & {}^{0}8 \\ \hline 0 & 0 & 9 & {}^{1}10 & {}^{2}11 & {}^{1}12 \\ \hline 0 & 0 & 13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & 28 \\ \hline 10 & 32 & 48 & 56 \\ \hline & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \\ \hline \end{array}$$

convolution filter

$$g = \begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

$$h = g * f = \begin{array}{|c|c|c|c|c|c|}\hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 2 & 3 & 4 \\ \hline {}^{1}0 & {}^{2}0 & {}^{1}5 & 6 & 7 & 8 \\ \hline {}^{0}0 & {}^{0}0 & {}^{0}9 & 10 & 11 & 12 \\ \hline {}^{1}0 & {}^{2}0 & {}^{1}13 & 14 & 15 & 16 \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline 1 & 4 & 8 & 12 \\ \hline 5 & 16 & 24 & 28 \\ \hline 10 & 32 & 48 & 56 \\ \hline 18 & & & \\ \hline \end{array}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$$f = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

convolution filter

$$g = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$h = g * f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 5 & 6 & 7 & 8 \\ 0 & 0 & 9 & 10 & 11 & 12 \\ 0 & 0 & 13 & 14 & 15 & 16 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 8 & 12 \\ 5 & 16 & 24 & 28 \\ 10 & 32 & 48 & 56 \\ 18 & 56 & & \end{bmatrix}$$

2D Convolution

a 2D grid of values that we want to convolve (e.g. an image)

$f =$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

convolution filter

$g =$

1	2	1
0	0	0
1	2	1

$h = g * f =$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	1	2	3	4
0	0	5	6	7	8
0	0	9	10	11	12
0	0	13	14	15	16

$=$

1	4	8	12
5	16	24	28
10	32	48	56
18	56	80	

2D Convolution

a 2D grid of values that we want
to convolve (e.g. an image)

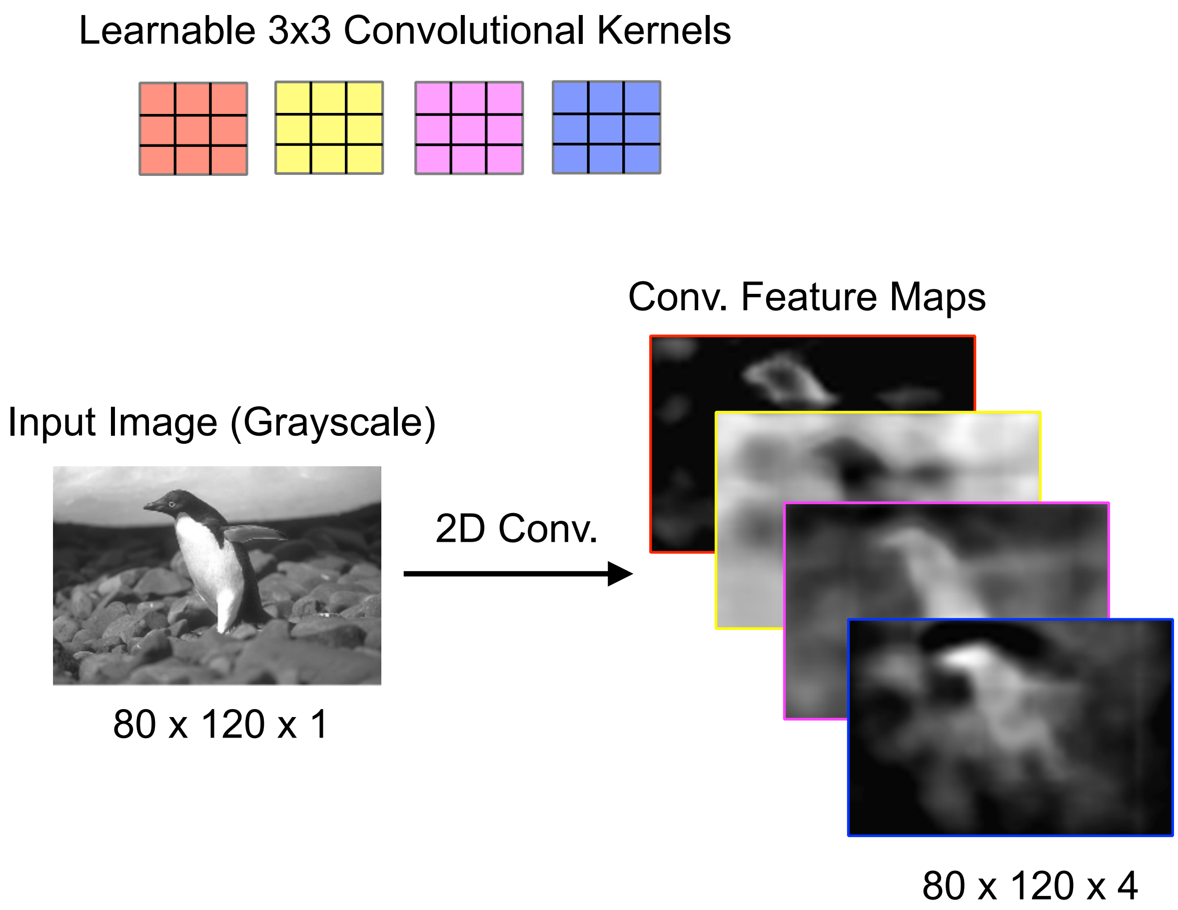
$$f = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

convolution filter

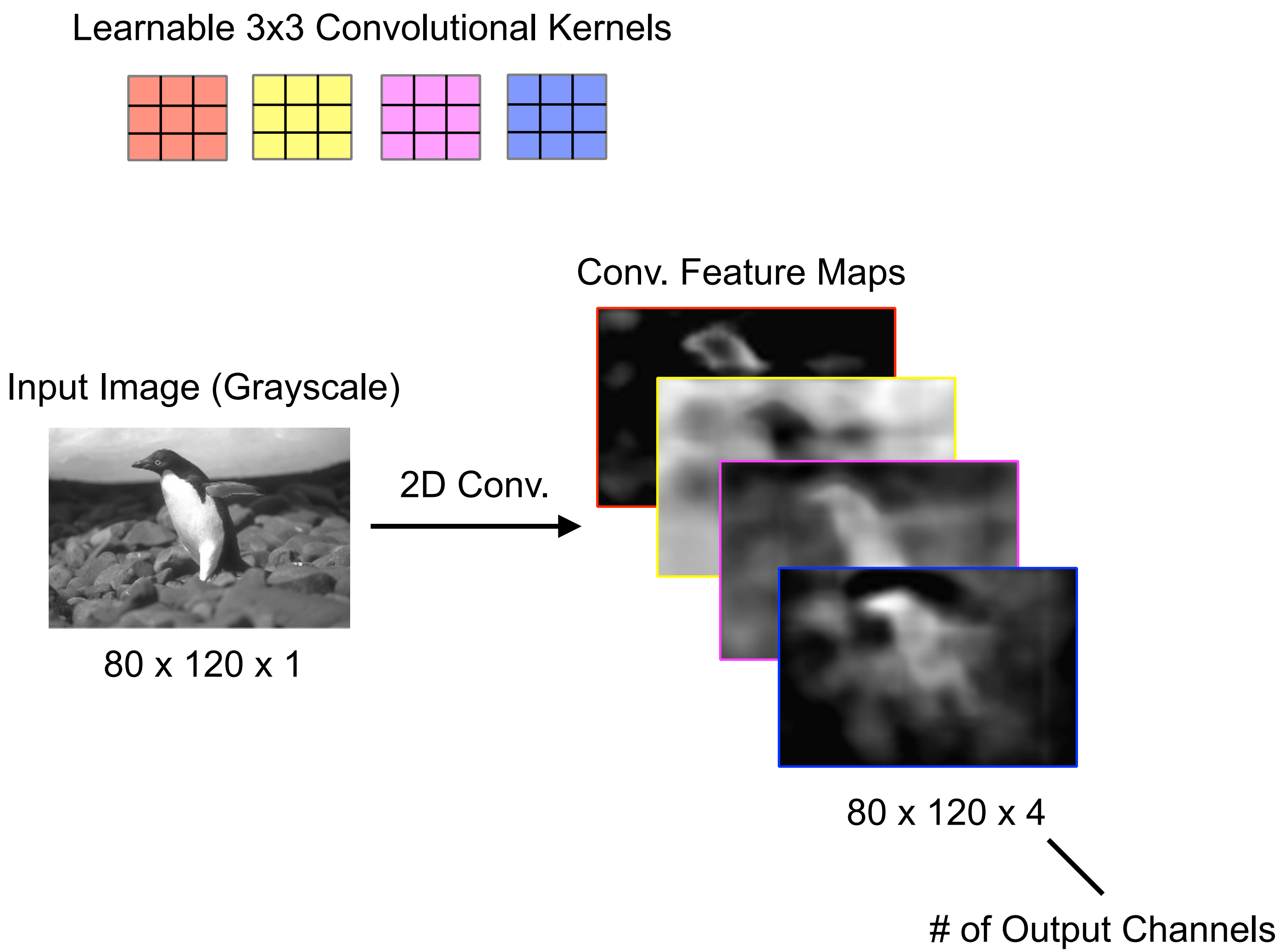
$$g = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$h = g * f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 5 & 6 & 7 & 8 \\ 0 & 0 & 9 & 10 & 11 & 12 \\ 0 & 0 & 13 & 14 & 15 & 16 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 8 & 12 \\ 5 & 16 & 24 & 28 \\ 10 & 32 & 48 & 56 \\ 18 & 56 & 80 & 88 \end{bmatrix}$$

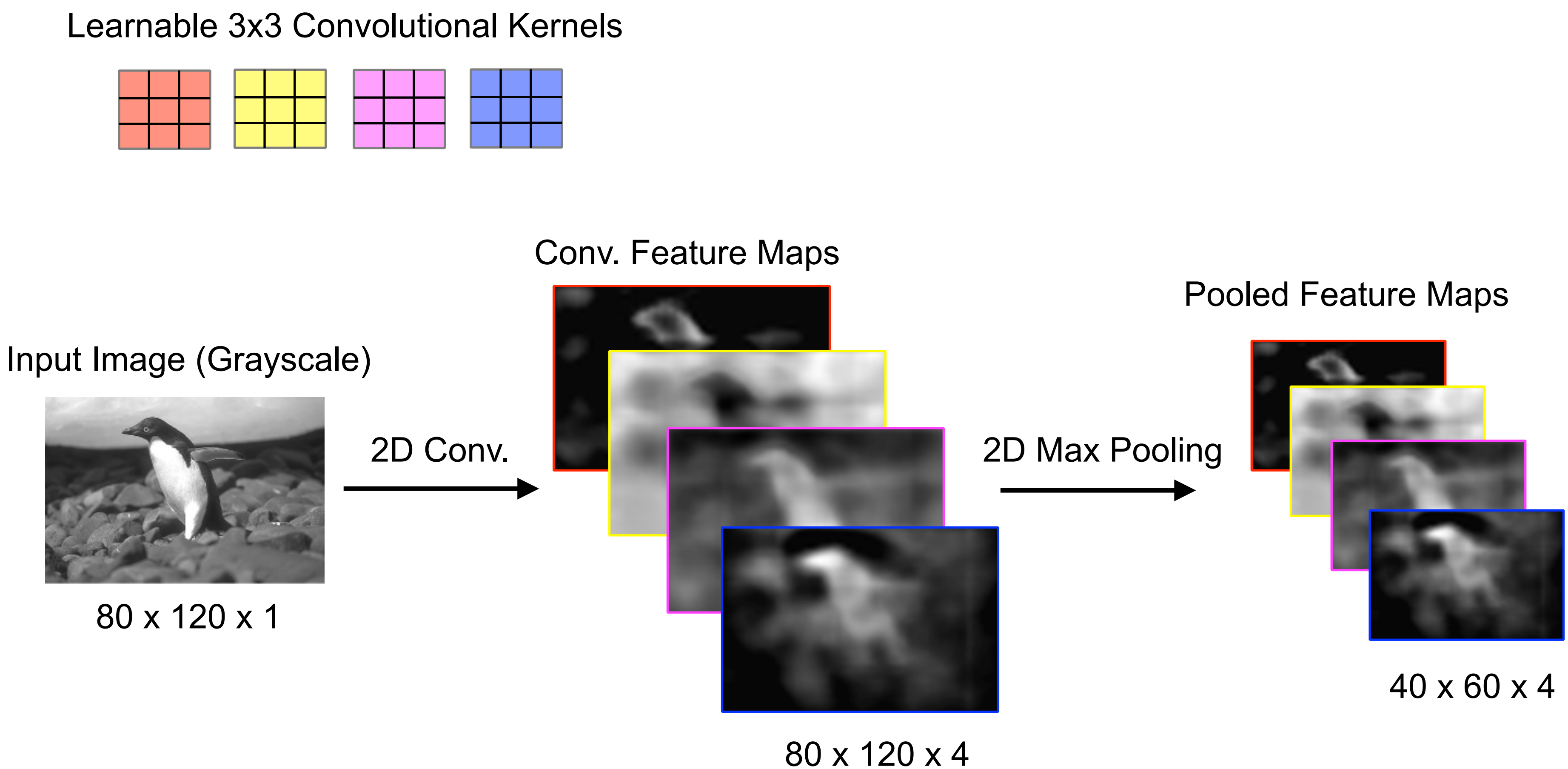
Convolutional Networks



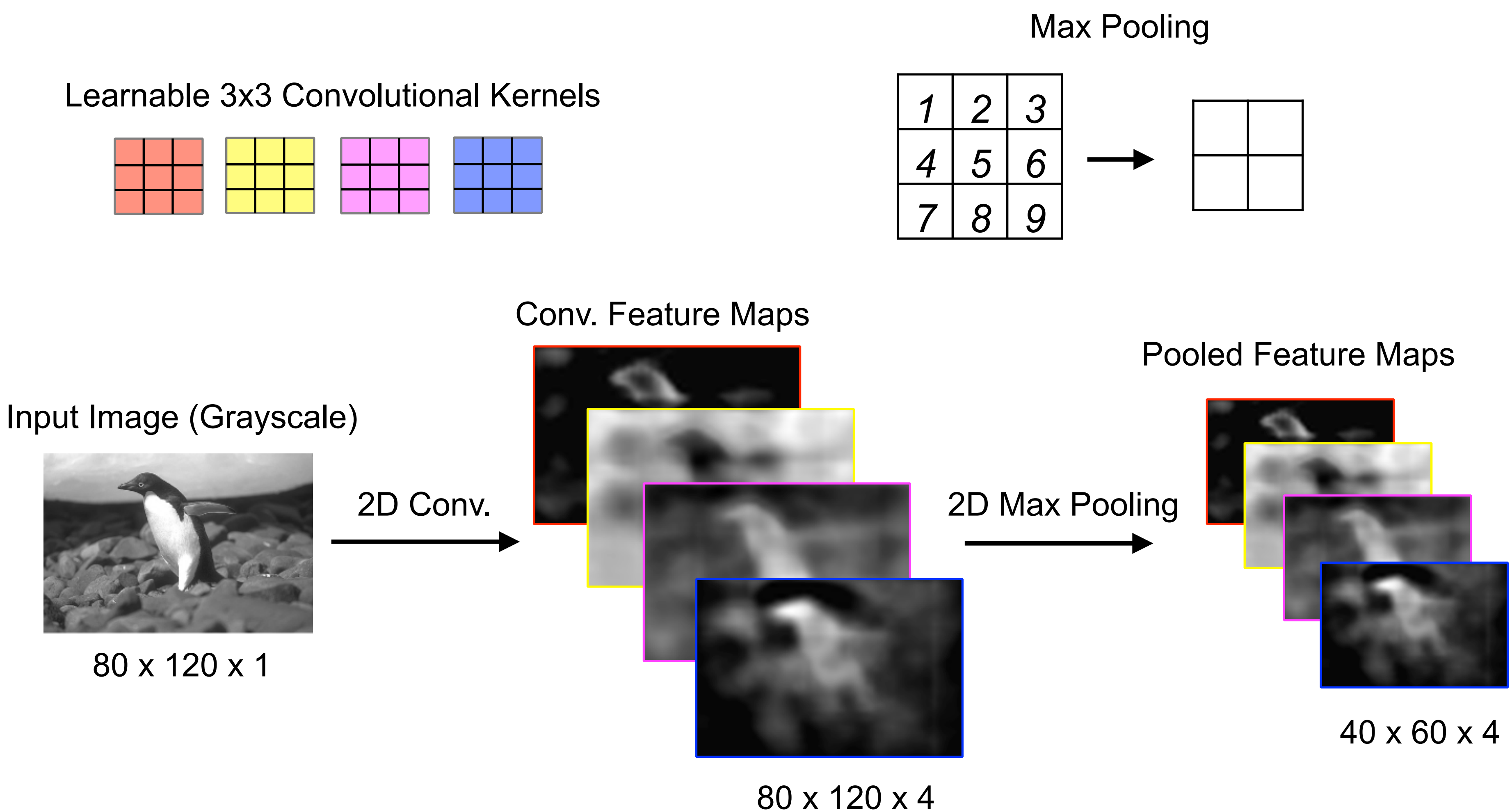
Convolutional Networks



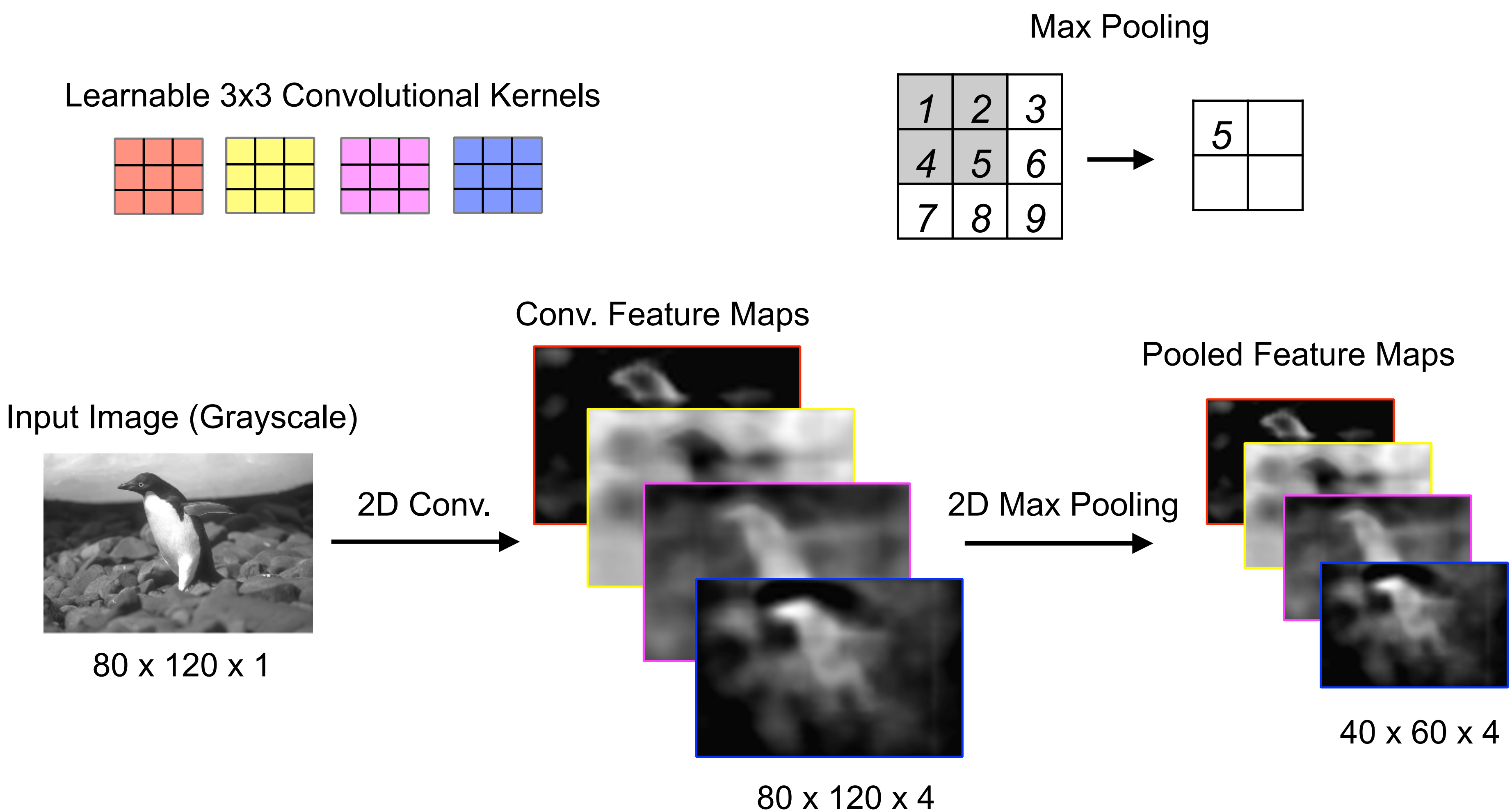
Convolutional Networks



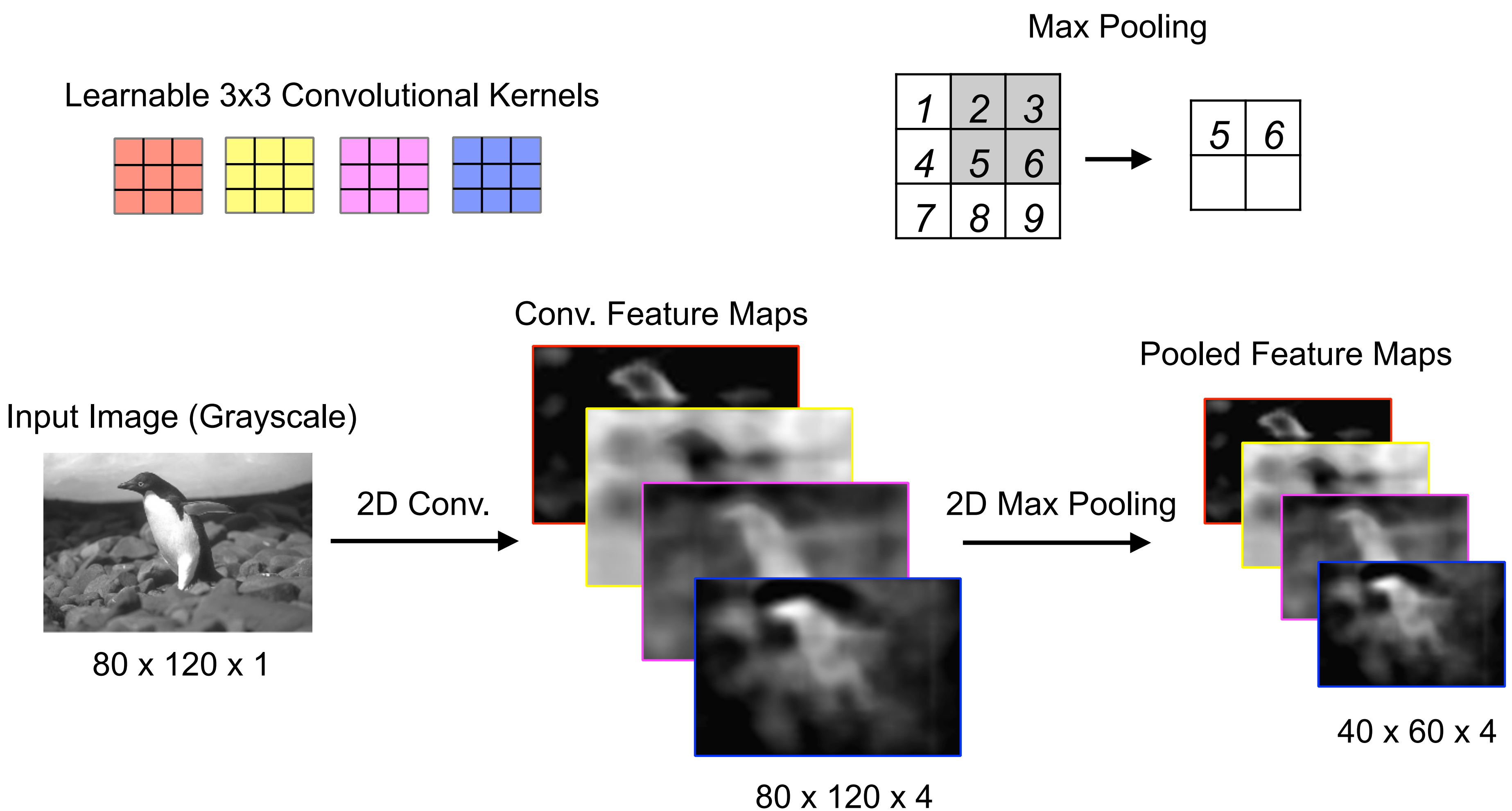
Convolutional Networks



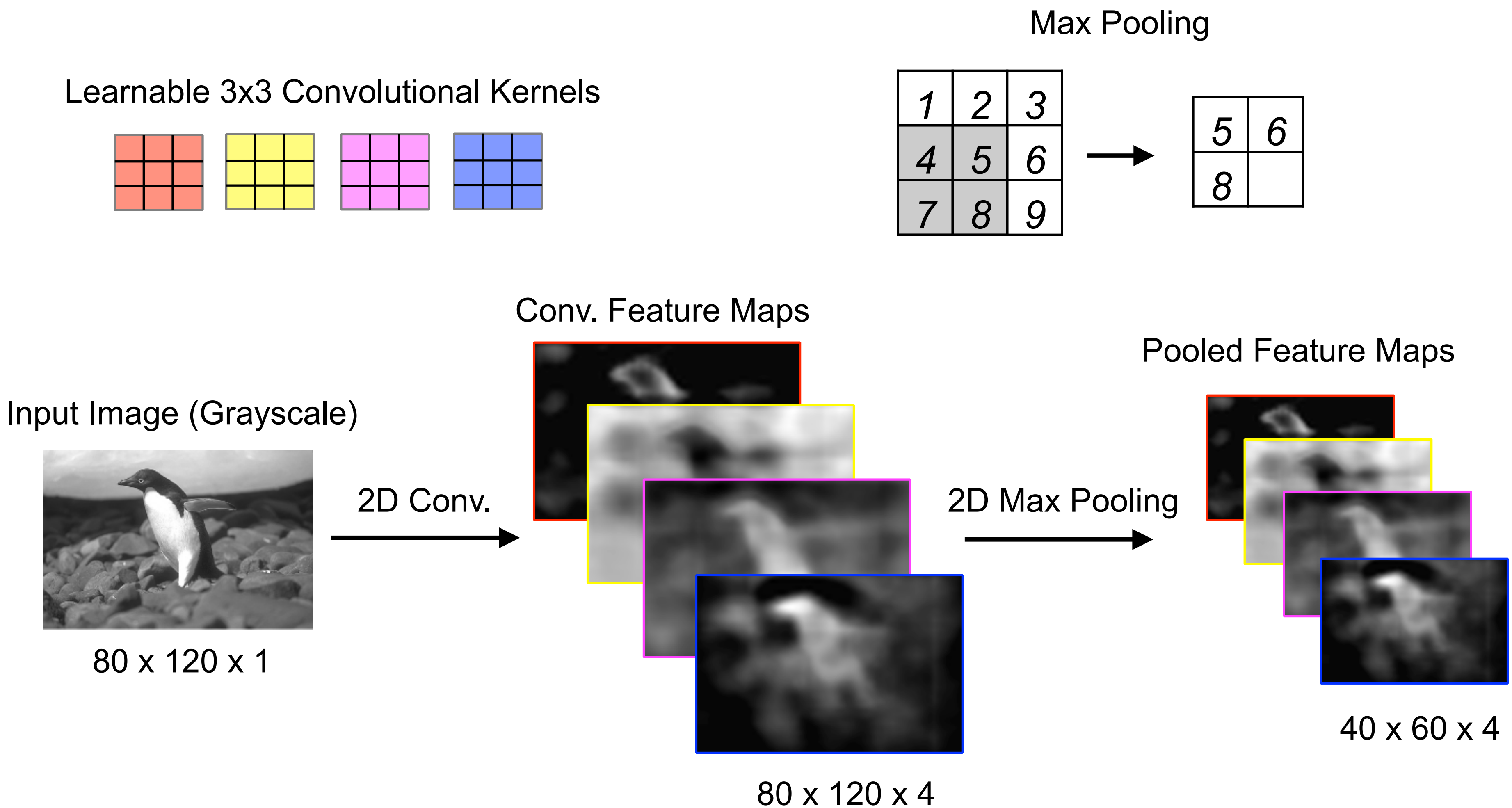
Convolutional Networks



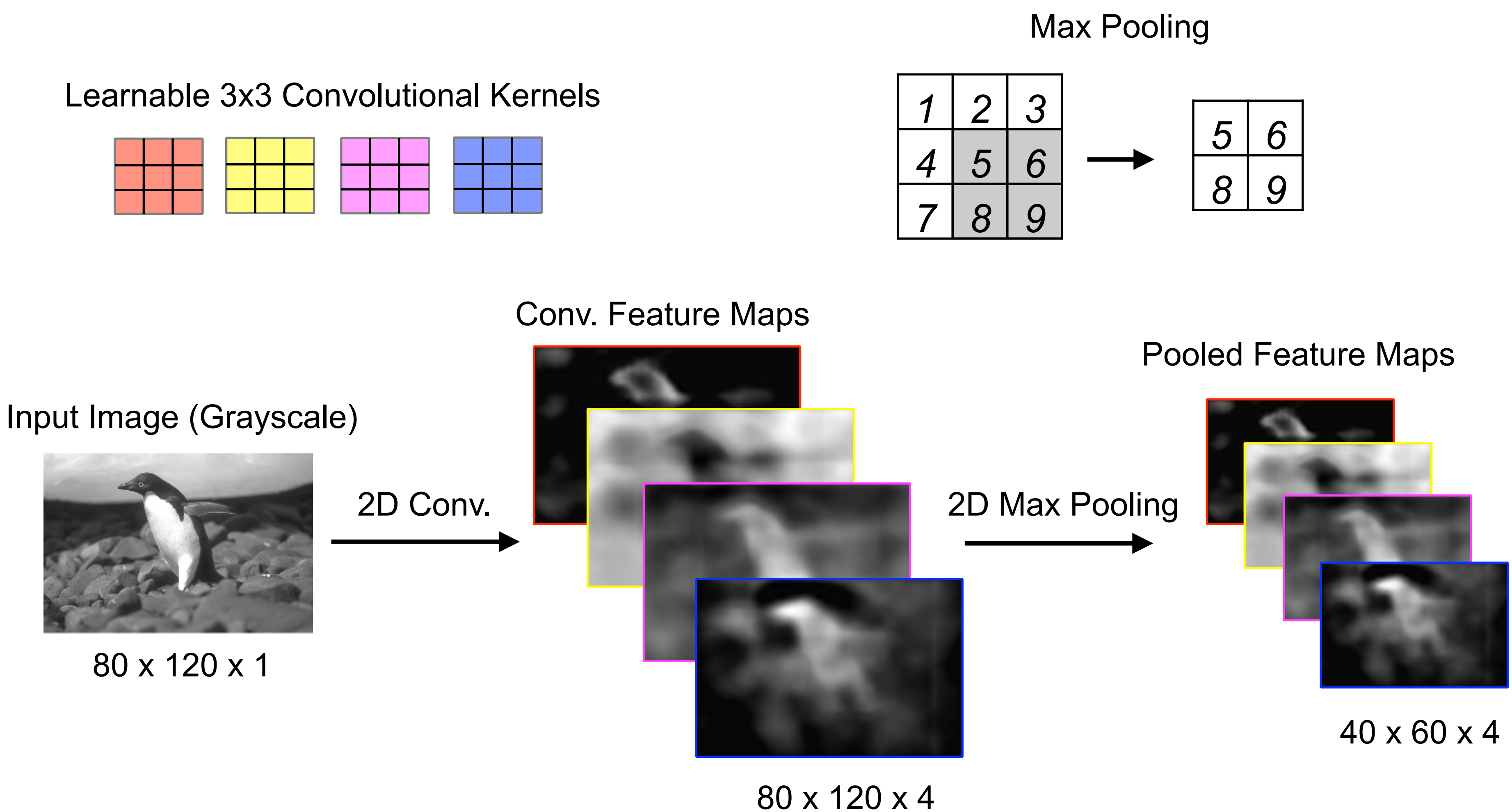
Convolutional Networks



Convolutional Networks

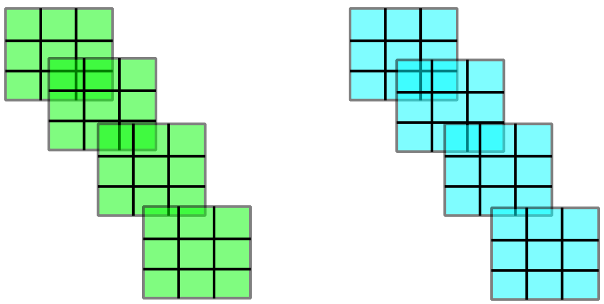


Convolutional Networks



Convolutional Networks

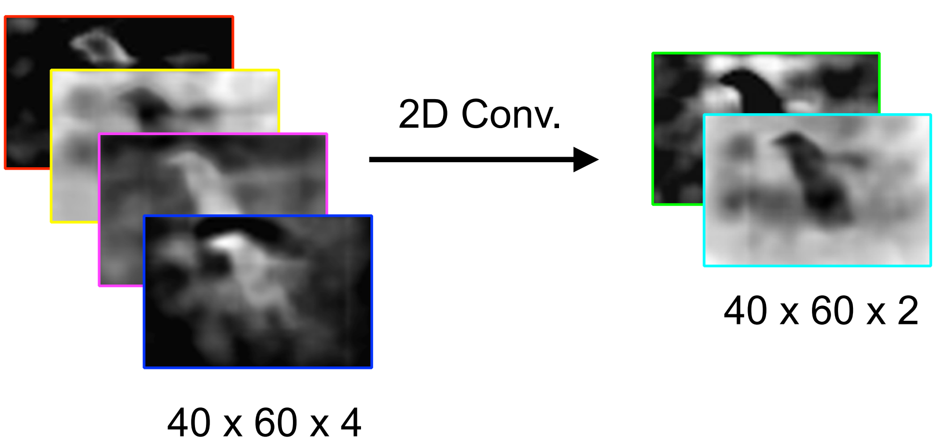
Learnable 4x3x3 Convolutional Kernels



2D Convolution on multi-channel inputs

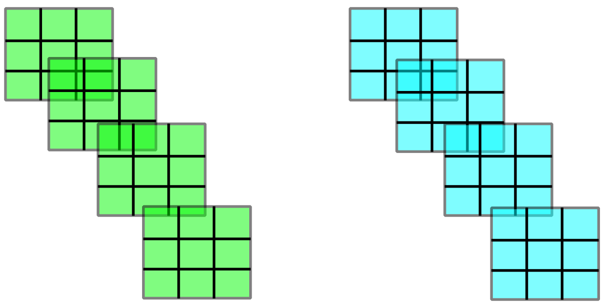
$$h(j) = \sum_{i}^{C_{in}} g(i, j) * f(i) \qquad j = 1, 2, ..., C_{out}$$

Pooled Feature Maps



Convolutional Networks

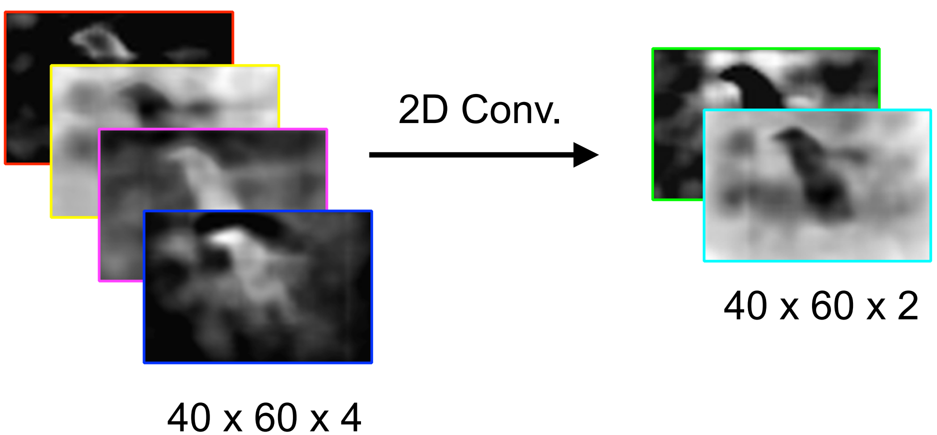
Learnable 4x3x3 Convolutional Kernels



2D Convolution on multi-channel inputs

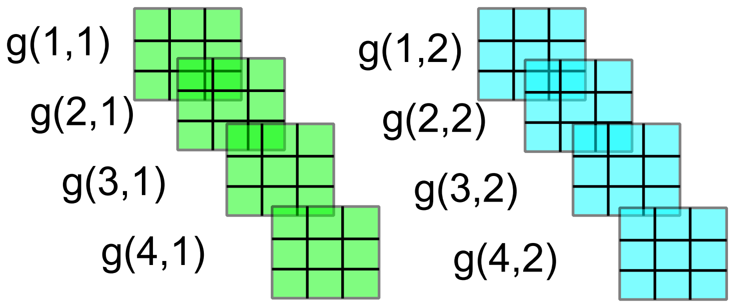
$$h(j) = \sum_{i}^{C_{in}} g(i,j) * f(i) \qquad j = 1, 2, ..., C_{out}$$

Pooled Feature Maps



Convolutional Networks

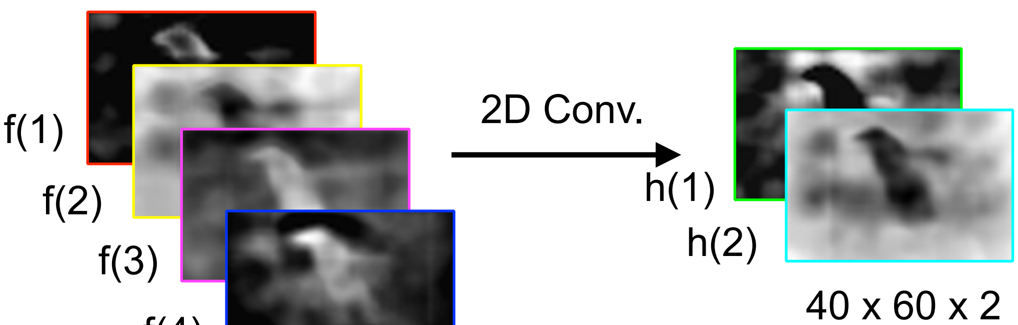
Learnable 4x3x3 Convolutional Kernels



2D Convolution on multi-channel inputs

$$\boxed{h(j) = \sum_{i}^{C_{in}} g(i,j) * f(i) \qquad j = 1, 2, ..., C_{out}}$$

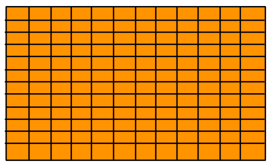
Pooled Feature Maps



40 x 60 x 4

Convolutional Networks

Learnable FC Layer Weight Matrix

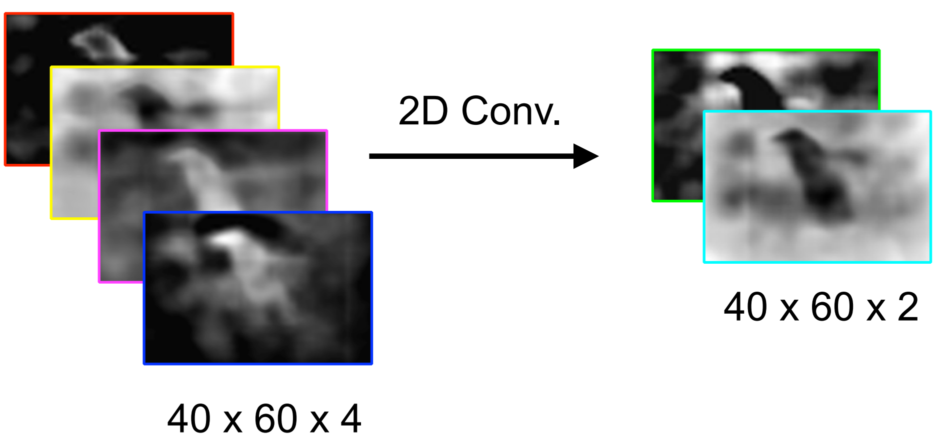


$$W \in \mathbb{R}^{d \times C}$$

d - feature dimensionality (4800 in this example)

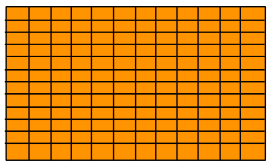
C - number of classes

Pooled Feature Maps



Convolutional Networks

Learnable FC Layer Weight Matrix

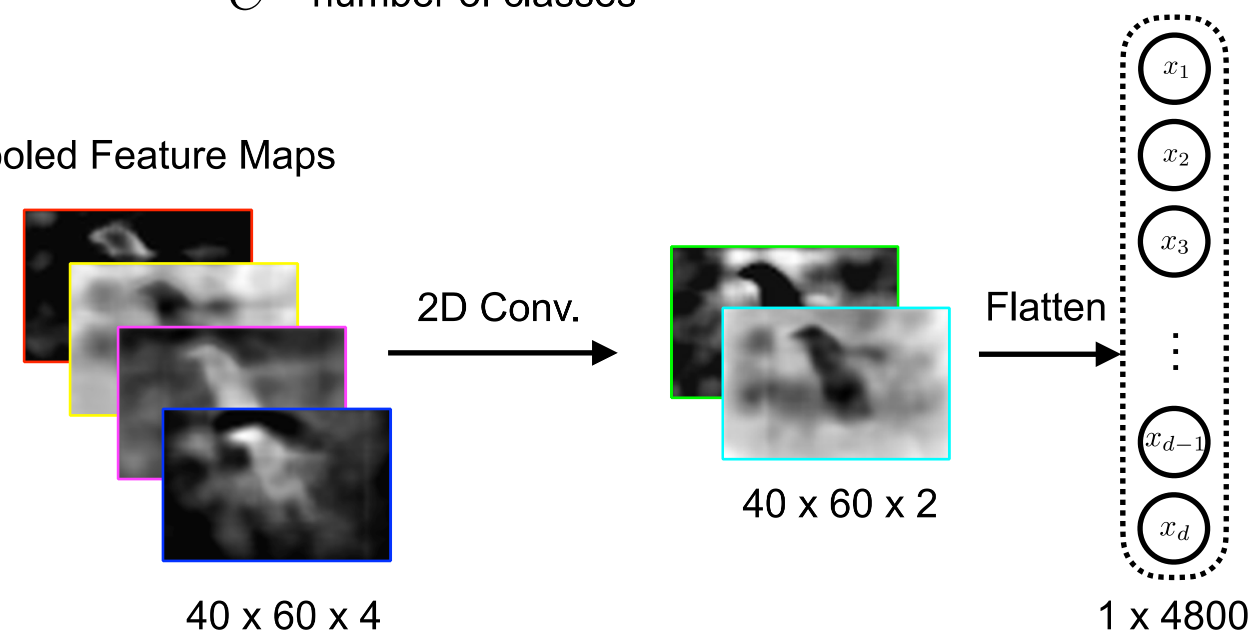


$$W \in \mathbb{R}^{d \times C}$$

d - feature dimensionality (4800 in this example)

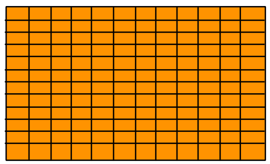
C - number of classes

Pooled Feature Maps



Convolutional Networks

Learnable FC Layer Weight Matrix

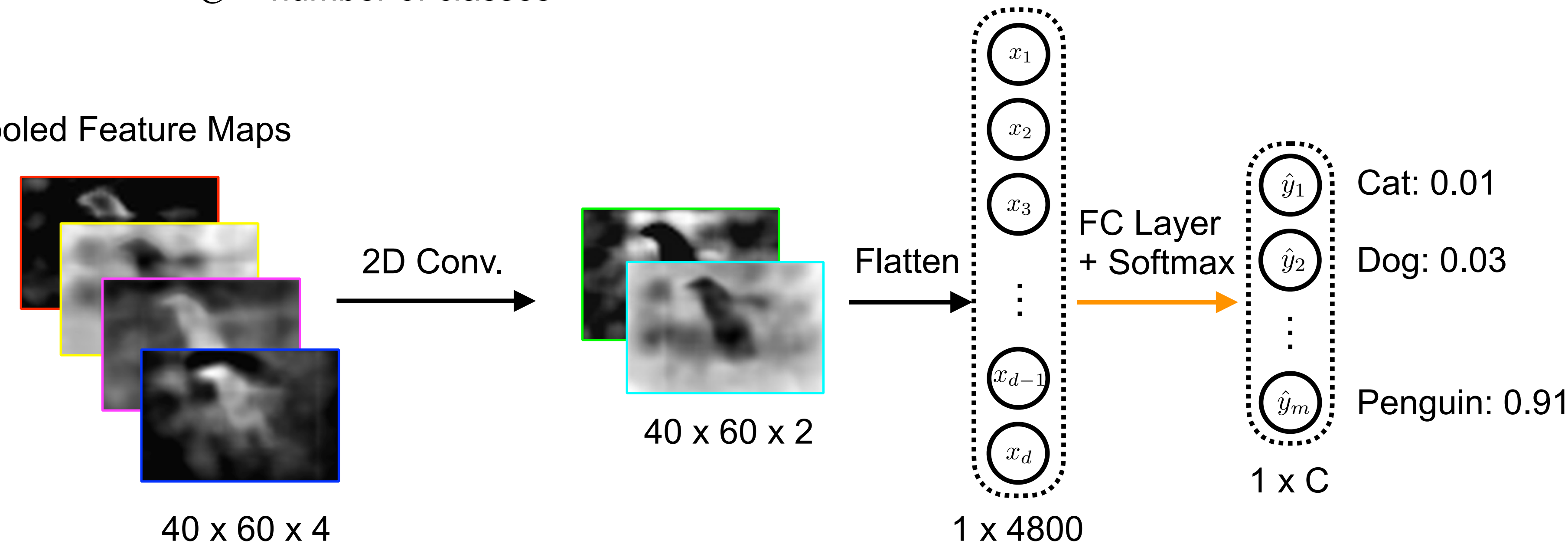


$$W \in \mathbb{R}^{d \times C}$$

d - feature dimensionality (4800 in this example)

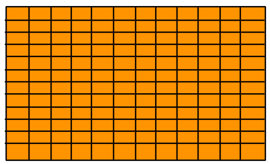
C - number of classes

Pooled Feature Maps



Convolutional Networks

Learnable FC Layer Weight Matrix



$$W \in \mathbb{R}^{d \times C}$$

d - feature dimensionality (4800 in this example)

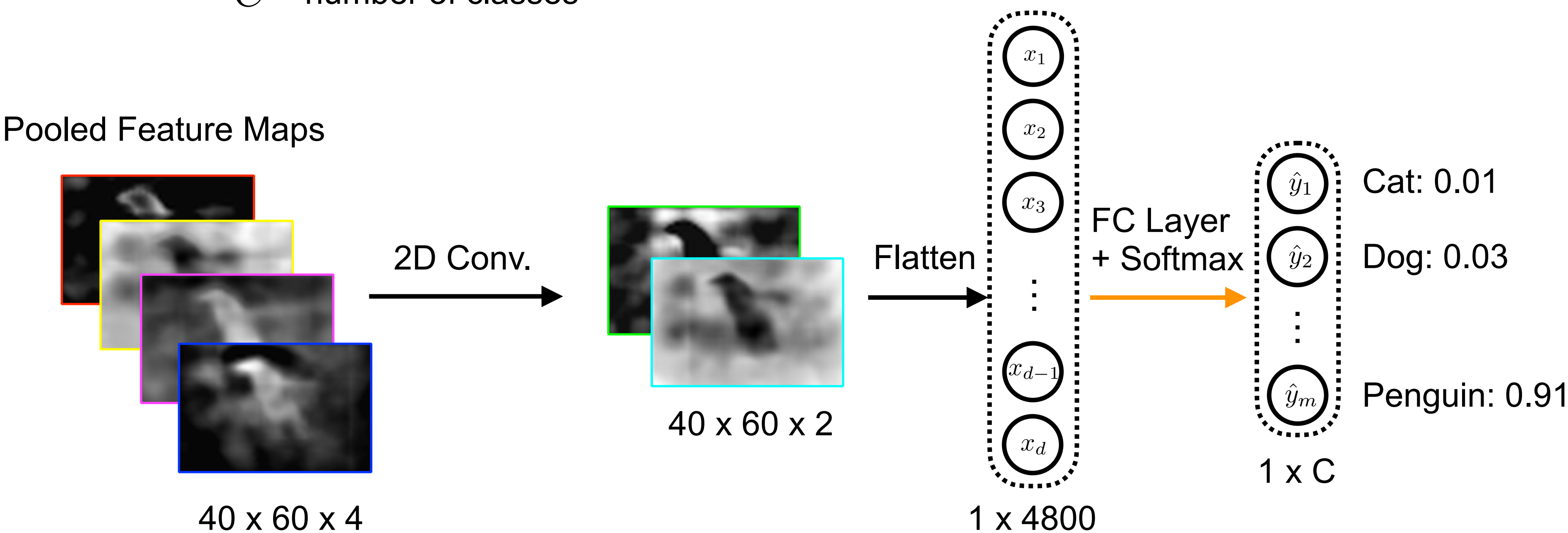
C - number of classes

Fully Connected Layers

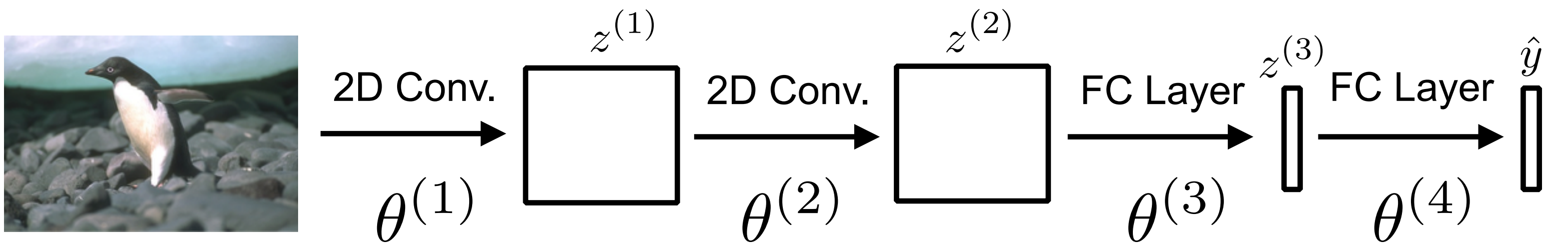
$$z = xW$$
$$\hat{y} = softmax(z)$$

$x \in \mathbb{R}^{1 \times d}$ - flattened feature vector

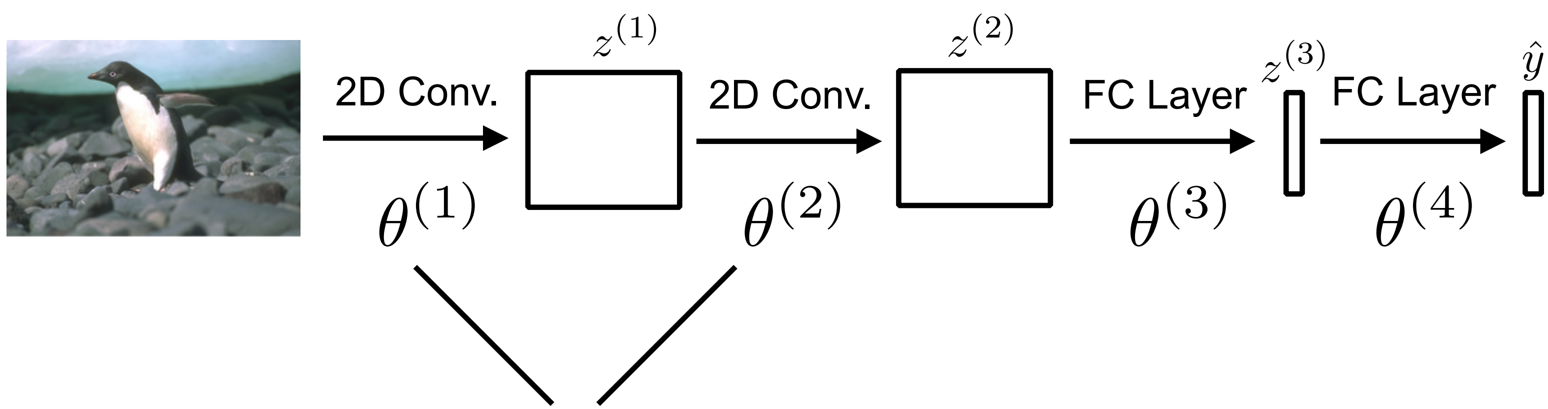
$\hat{y} \in \mathbb{R}^{1 \times C}$ - predicted probabilities



Convolutional Networks

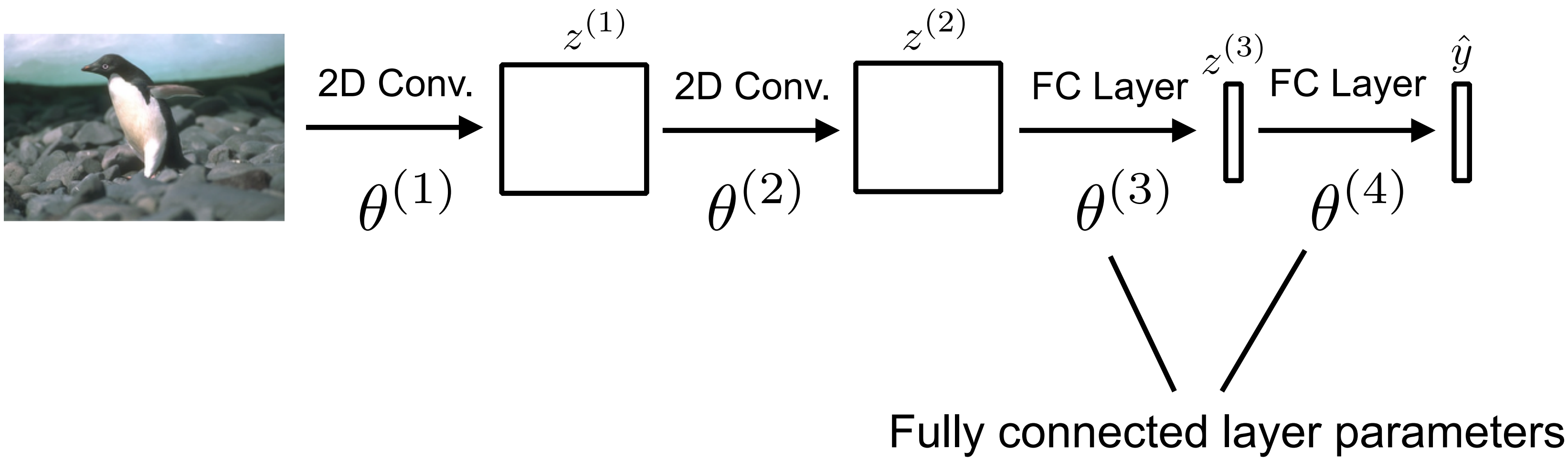


Convolutional Networks

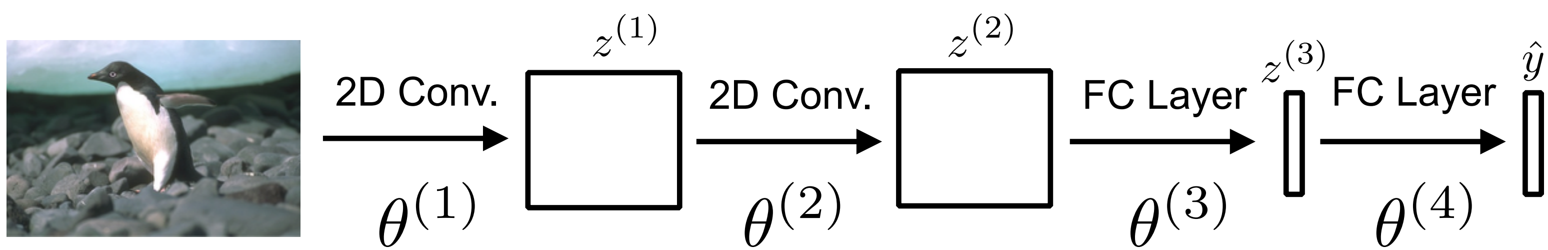


Convolutional layer parameters in layers 1 and 2

Convolutional Networks

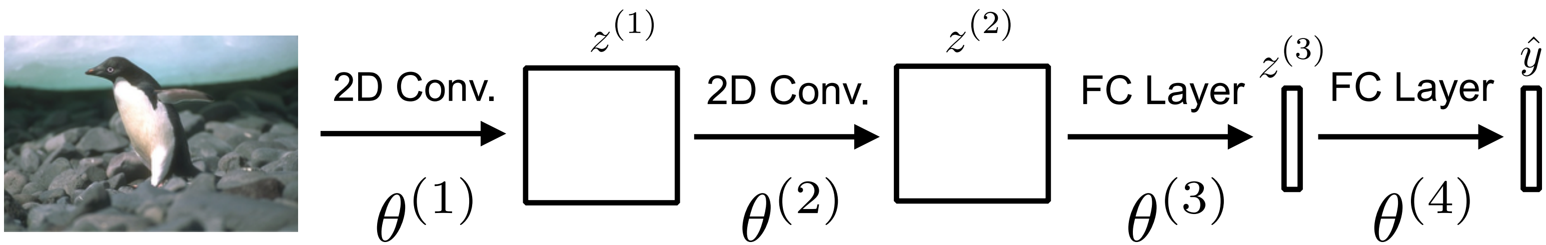


Convolutional Networks



How do we learn these parameters from the data?

Convolutional Networks



- Assume that we are a given a **labeled** training dataset

$$\{(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\}$$

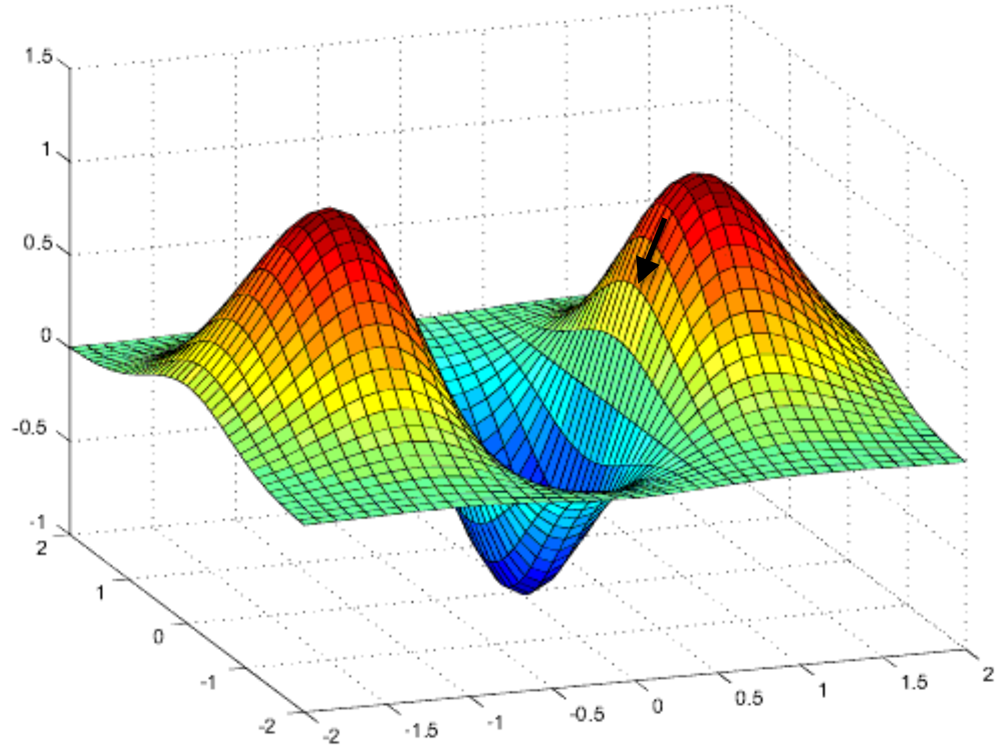
- We want to adjust the parameters of the network such that its predictions would match true labels.

Backpropagation

Gradient descent:

- Iteratively minimizes the objective function.
- The function needs to be differentiable.

$$\theta = \theta - \alpha \frac{\partial L(\theta)}{\partial \theta}$$

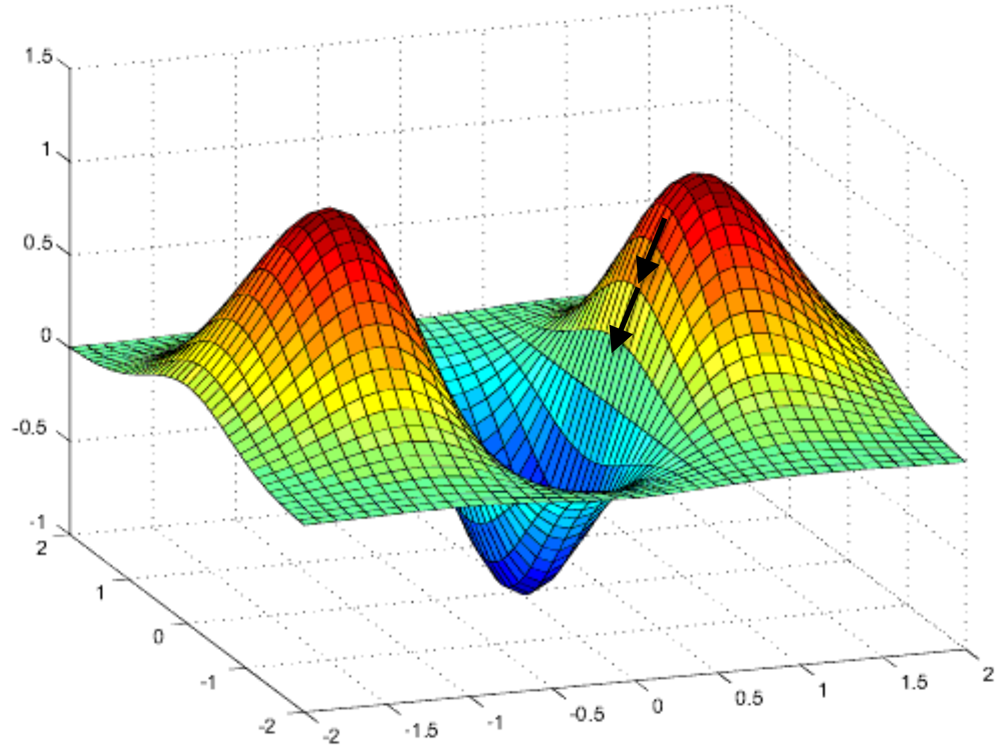


Backpropagation

Gradient descent:

- Iteratively minimizes the objective function.
- The function needs to be differentiable.

$$\theta = \theta - \alpha \frac{\partial L(\theta)}{\partial \theta}$$

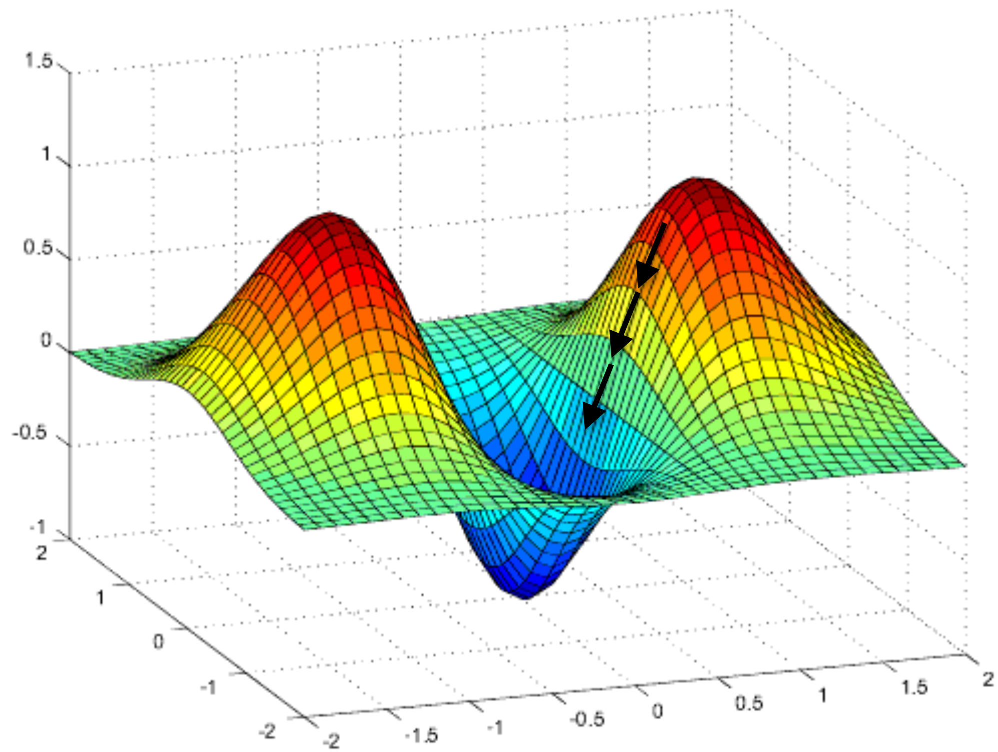


Backpropagation

Gradient descent:

- Iteratively minimizes the objective function.
- The function needs to be differentiable.

$$\theta = \theta - \alpha \frac{\partial L(\theta)}{\partial \theta}$$

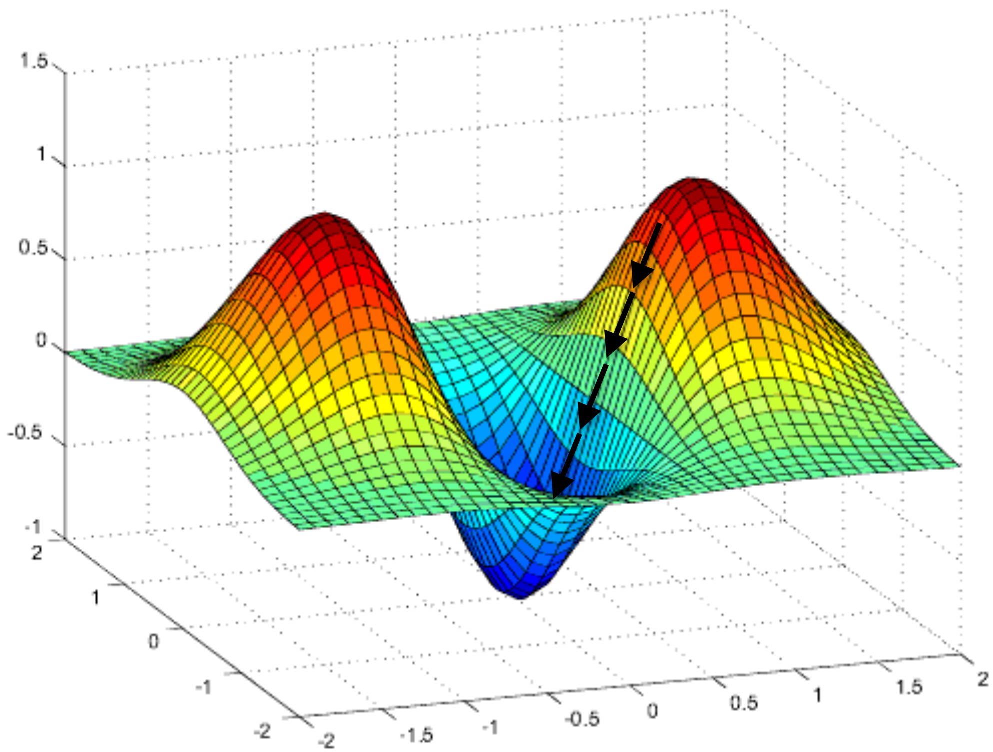


Backpropagation

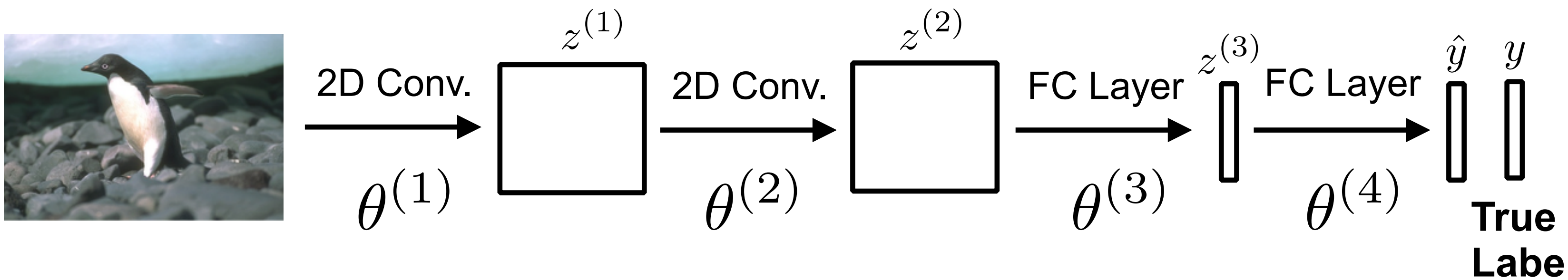
Gradient descent:

- Iteratively minimizes the objective function.
- The function needs to be differentiable.

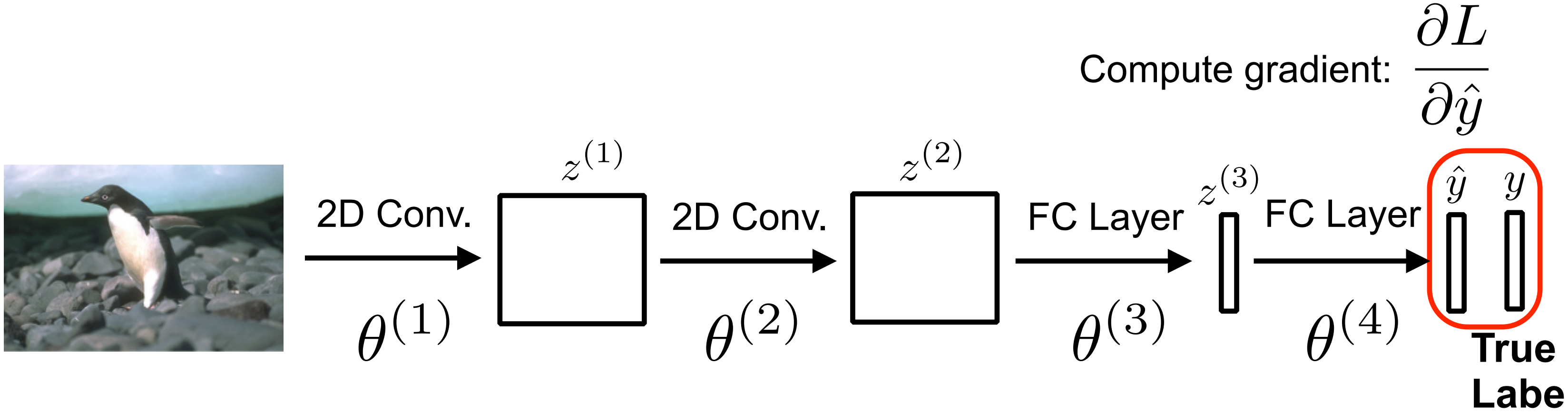
$$\theta = \theta - \alpha \frac{\partial L(\theta)}{\partial \theta}$$



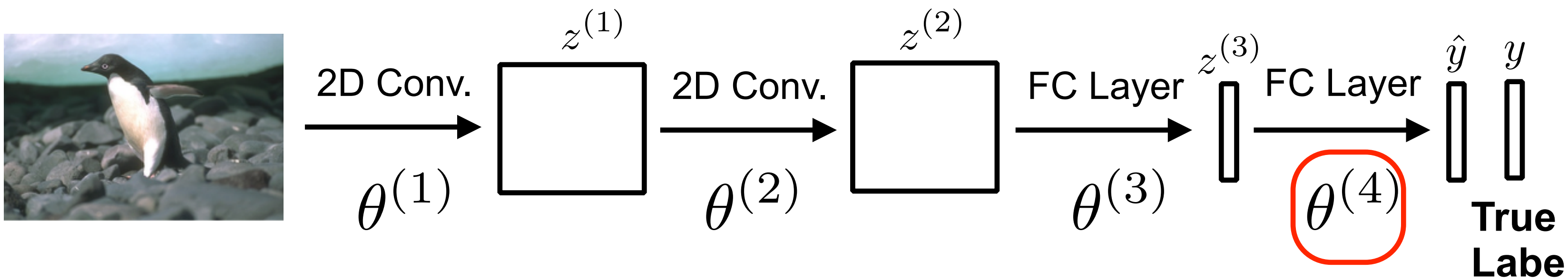
Backpropagation



Backpropagation



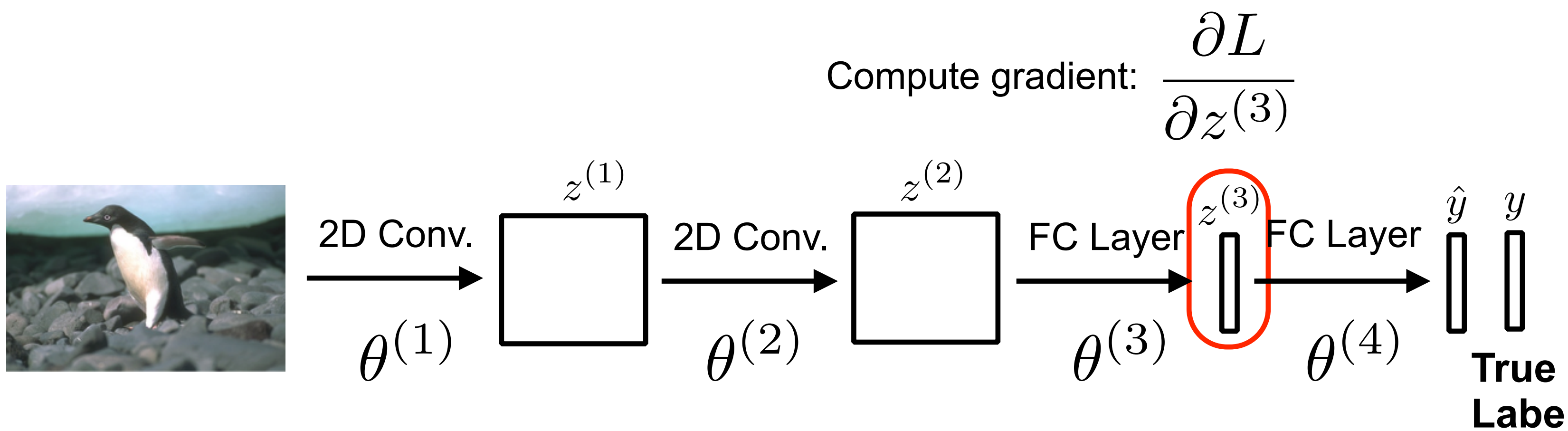
Backpropagation



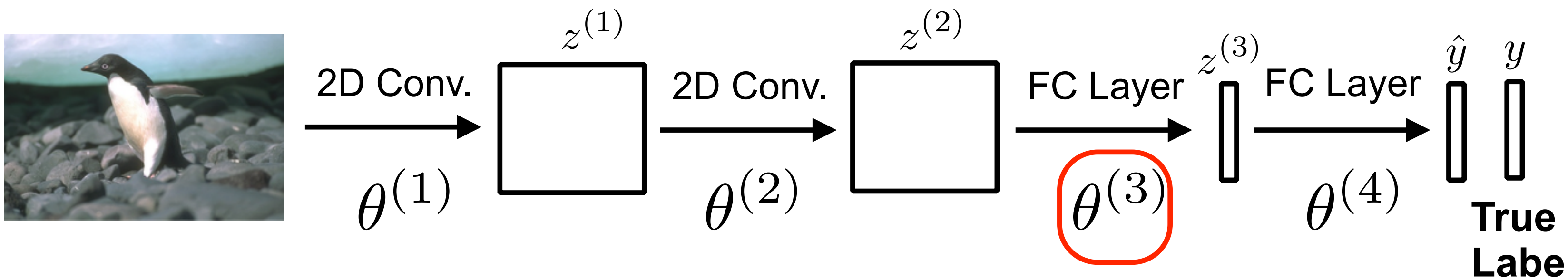
Compute gradient: $\dfrac{\partial L}{\partial \theta^{(4)}}$

Update parameters: $\theta^{(4)} = \theta^{(4)} - \alpha \dfrac{\partial L}{\partial \theta^{(4)}}$

Backpropagation



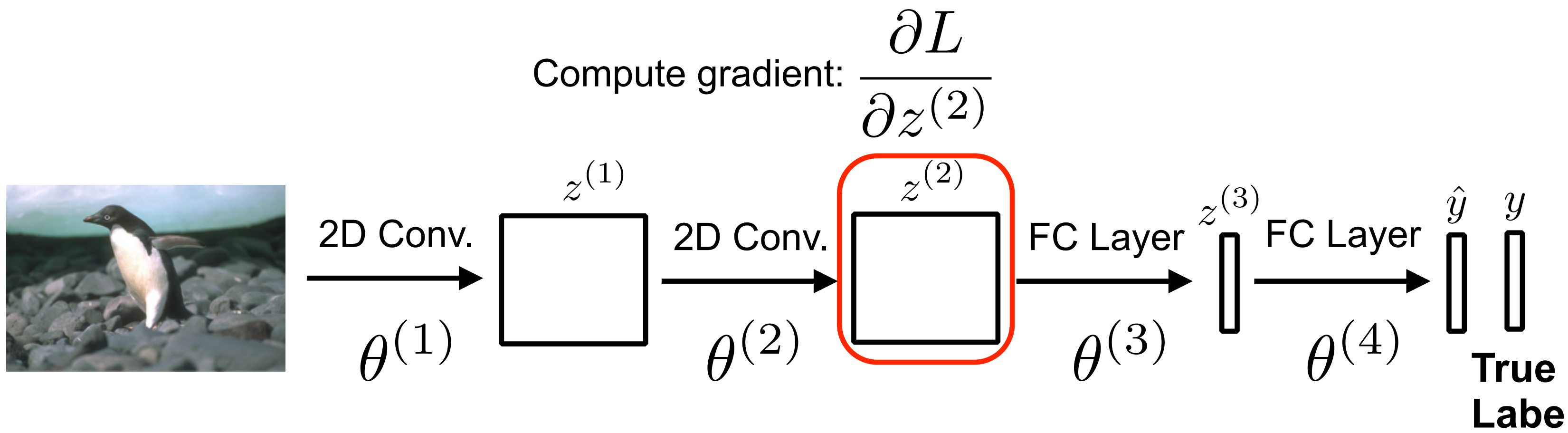
Backpropagation



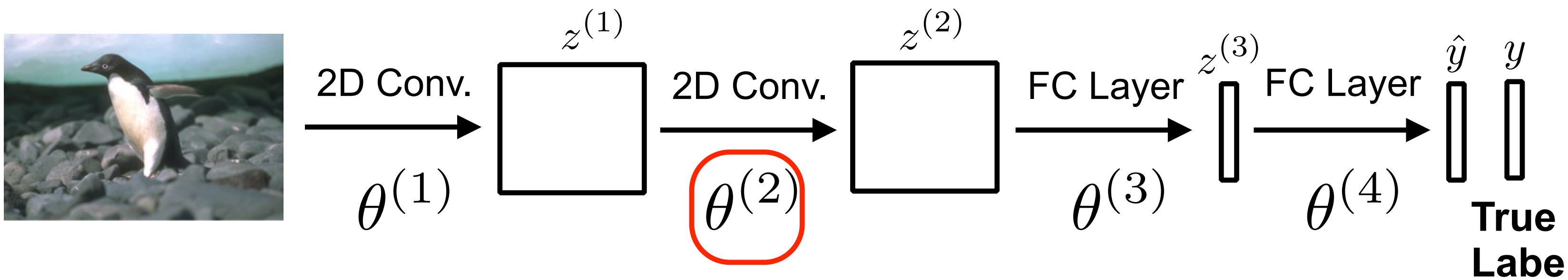
Compute gradient: $\dfrac{\partial L}{\partial \theta^{(3)}}$

Update parameters: $\theta^{(3)} = \theta^{(3)} - \alpha \dfrac{\partial L}{\partial \theta^{(3)}}$

Backpropagation



Backpropagation

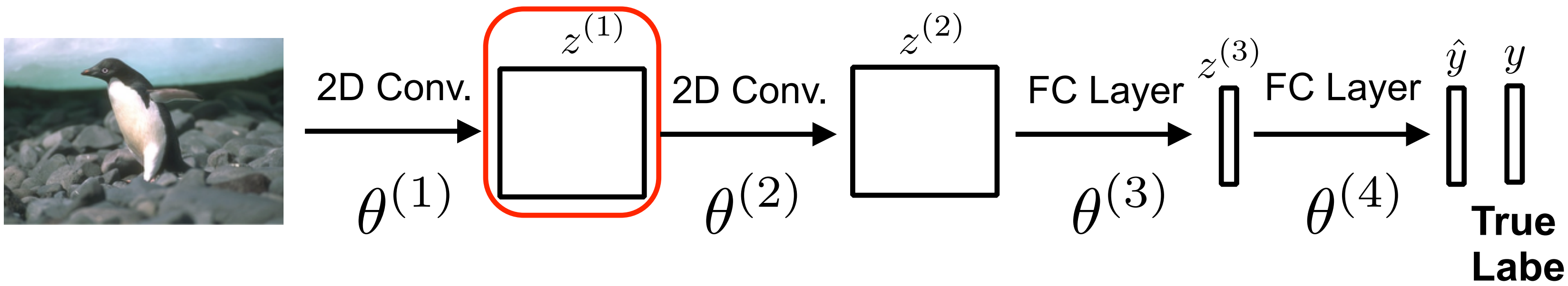


Compute gradient: $\dfrac{\partial L}{\partial \theta^{(2)}}$

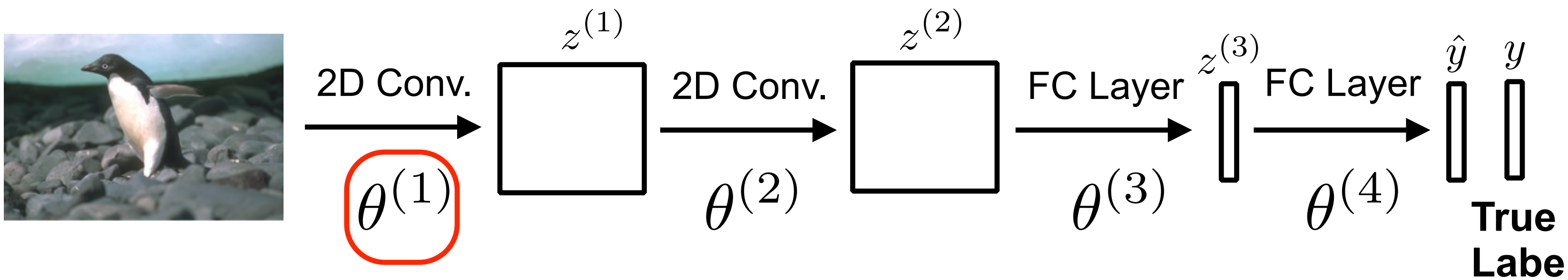
Update parameters: $\theta^{(2)} = \theta^{(2)} - \alpha \dfrac{\partial L}{\partial \theta^{(2)}}$

Backpropagation

Compute gradient: $\dfrac{\partial L}{\partial z^{(1)}}$



Backpropagation



Compute gradient: $\dfrac{\partial L}{\partial \theta^{(1)}}$

Update parameters: $\theta^{(1)} = \theta^{(1)} - \alpha \dfrac{\partial L}{\partial \theta^{(1)}}$

Backpropagation

1. Let $\frac{\partial L}{\partial z^{(n)}} = \hat{y} - y$ where n is the number of layers in the network.

2. For each layer l:

   a) Compute the gradients: $\frac{\partial L}{\partial z^{(l)}} = \frac{\partial L}{\partial z^{(l+1)}} \frac{\partial z^{(l+1)}}{\partial \sigma\left(z^{(l)}\right)} \frac{\partial \sigma\left(z^{(l)}\right)}{\partial z^{(l)}}$

   where σ is a non-linear function (e.g., ReLU).

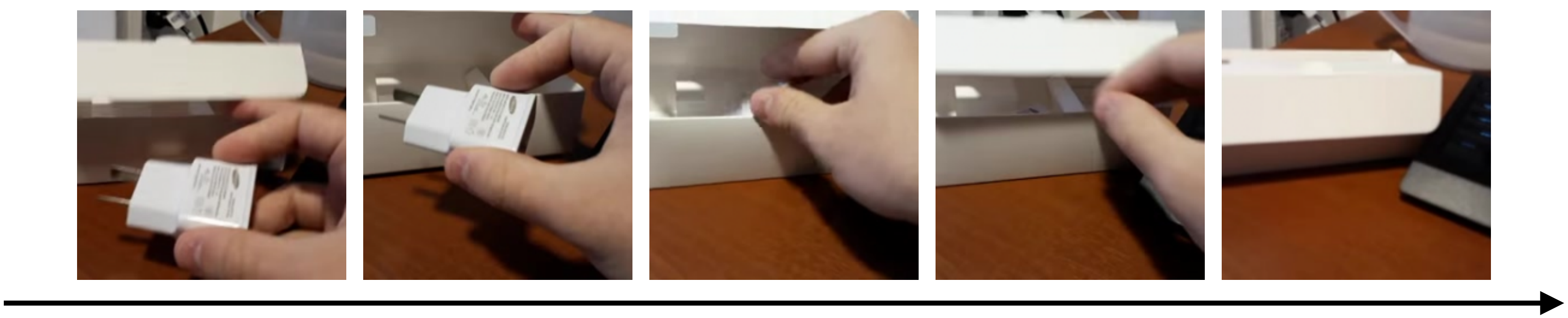
   b) Compute partial derivatives: $\frac{\partial L}{\partial \theta^{(l)}} = \frac{\partial L}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial \theta^{(l)}}$

   c) Update the parameters: $\theta^{(l)} = \theta^{(l)} - \alpha \frac{\partial L}{\partial \theta^{(l)}}$

Video Modeling

How can we apply CNNs to video for modeling temporal information?

Input Video



time

First Assignment

- The reading list is posted here.

- Select the following:
    1. Seven 30min or 45min papers for standard paper presentations (marked **red** and **purple** in the schedule). Any combo of the papers suffice (e.g., five 30min & two 45min papers, all 30min papers, etc.)
    2. Three 20min papers for paper battles (marked **green** in the schedule).

- Make sure that the papers that you selected will **NOT** be presented by me.

- Rank the papers in each of these lists in descending order of preference (from highest to lowest) and upload them to Canvas **by Sunday, Aug 27th, 11:59 PM** (please include paper IDs in your lists!!).

- I will then update the website with the paper assignments.

Second Assignment

- Complete the paper critique for paper [5] SlowFast Networks for Video Recognition.

- Upload it to Canvas by **1 PM on Wednesday, August 30th**.