ICARC

International Conference on Advanced Research in Computing

Curriculum Learning
An Efficient Learning Paradigm

SLIIT UNI
THE KNOWLEDGE UNIVERSITY

Utrecht University

SLIIT Research and International

BRAIN LABS
Brain-Inspired AI & Neuroinformatics Research Group

Organizing Committee

Dr. Dharshana Kasthurirathna
Sri Lanka Institute of Information Technology (SLIIT)

Dr. Mahima Weerasinghe
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Menan Velayuthan
Utrecht University

Mr. Asiri Gawesha
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Sanka Mohottala
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Dulara Madhusanka
Sri Lanka Institute of Information Technology (SLIIT)

Ms. Savini Kommalage
Sri Lanka Institute of Information Technology (SLIIT)

Organized by BrAINLabs Research Group, SLIIT
Funded by a SLIIT Research & International (Grant No. PVC(R&I)/RG/2025/12)

SLIIT UNI THE KNOWLEDGE UNIVERSITY
Utrecht University
ICARC
BRAIN LABS

Resource Personal

Ms. Savini Kommalage
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Dulara Madhusanka
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Menan Velayuthan
Utrecht University

Mr. Asiri Gawesha
Sri Lanka Institute of Information Technology (SLIIT)

Mr. Sanka Mohottala
Sri Lanka Institute of Information Technology (SLIIT)

Curriculum Learning
An Efficient Learning Paradigm

Session 3 : Applying pre-defined curriculum learning to an image classification problem

Savini Kommalage
Sri Lanka Institute of Information Technology (SLIIT)

Scan to follow along

Google Collab Code

ICARC Tutorial Website

Shape Classification Using Curriculum Learning [1]

- Task : 3-class image classification (rectangle, ellipse, triangle)

- Input : 32 × 32 grayscale images
Basic Shapes — low variability (easy)
Geometric Shapes — high variability (hard)

- Model : Neural network architecture, Stochastic Gradient Descent (SGD)

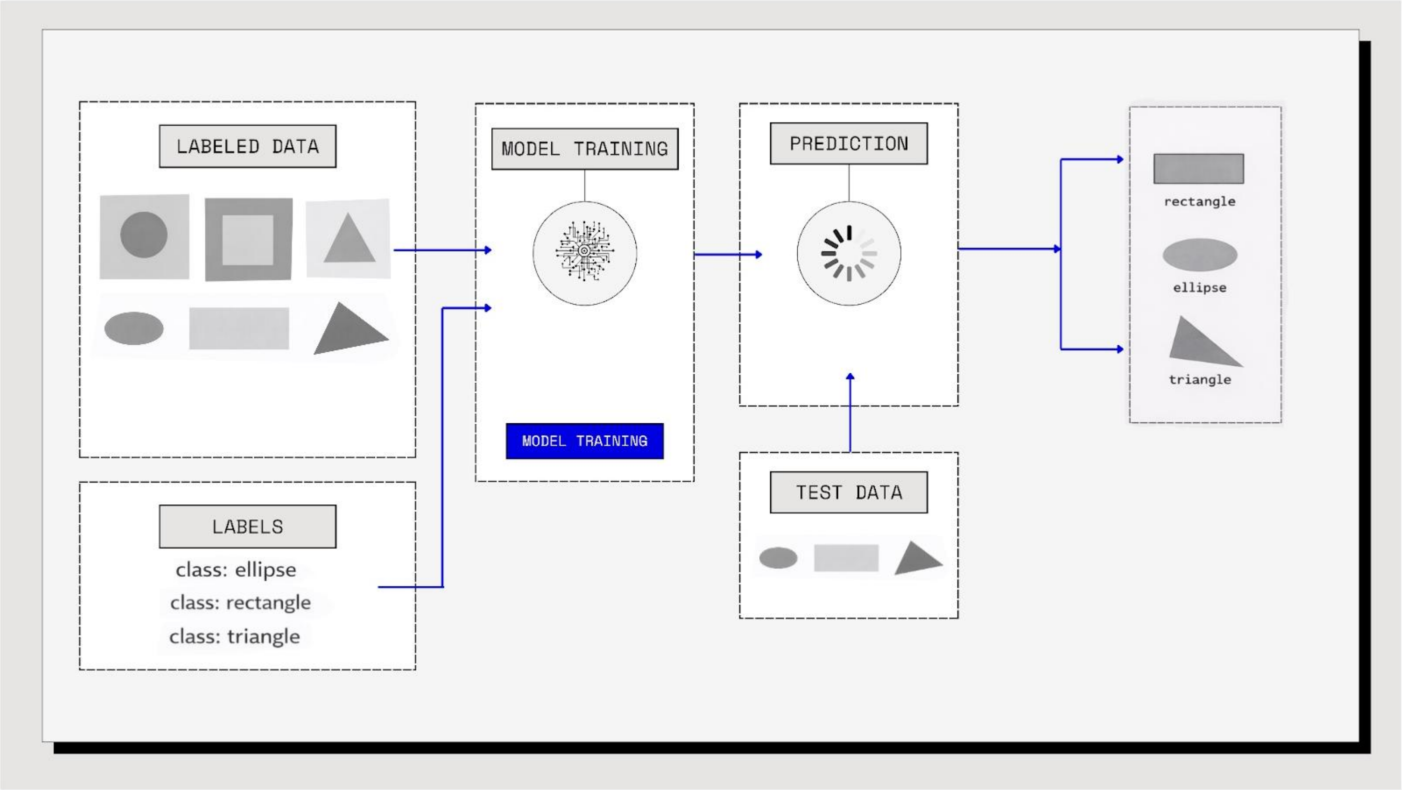


Figure : Experimental set up :Two-stage curriculum design for synthetic shape classification

Shape Classification Dataset

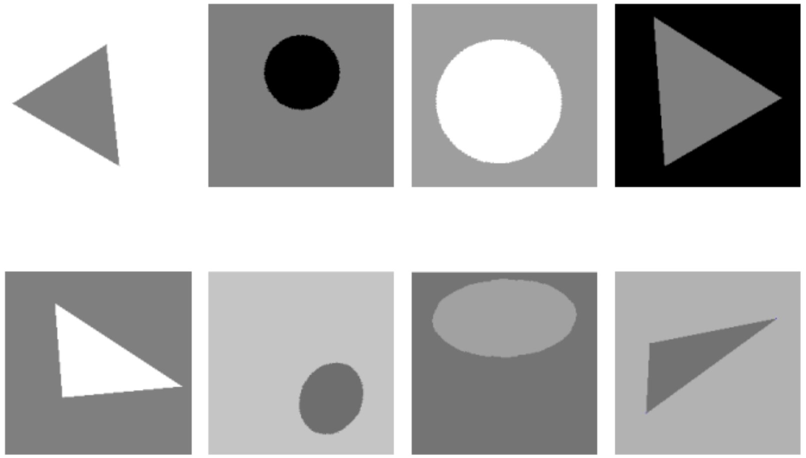


Figure : “Images are shown here with a higher resolution than the actual dataset (32x32 pixels)” - Curriculum Learning (Bengio et al.)

The original synthetic dataset used in the paper Curriculum Learning by Bengio et al. is not publicly available. Therefore, we reconstruct the dataset as faithfully as possible based on the methodological description provided in the paper.

Image resolution: 32 × 32 pixels
Grayscale images

BasicShapes (low variability)

- circle
- square (equal sides, no rotation)
- equilateral triangle (no rotation)

GeomShapes (full variability)

- ellipse with varying aspect ratio + rotation
- rectangle with varying width/height + rotation
- triangle with randomly sampled vertices + rotation

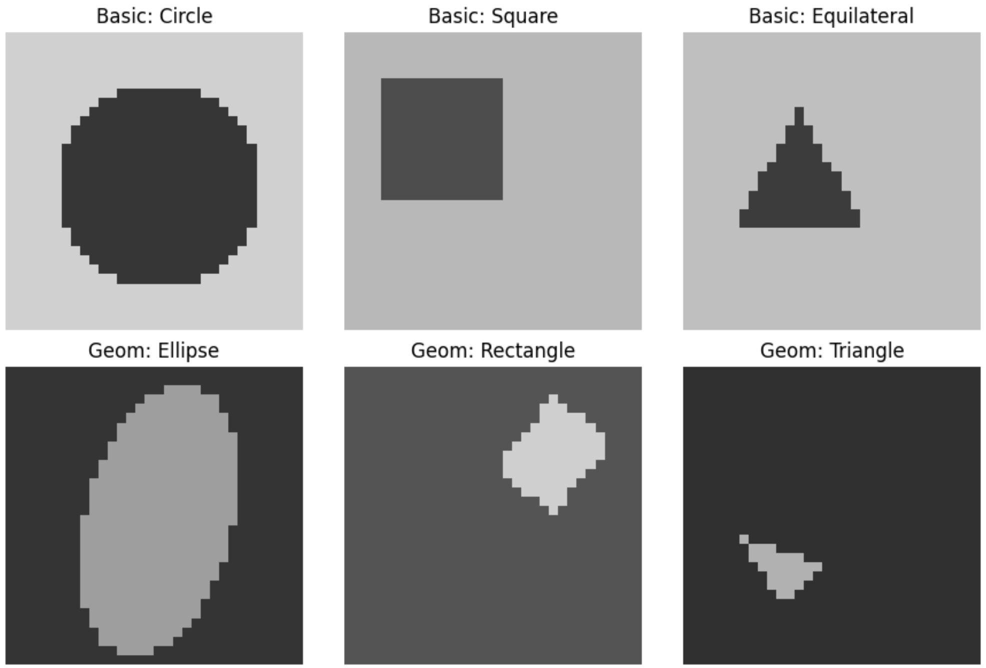


Figure : synthetic dataset

Switch epoch pacing

Experiment conducted over 20 different random seeds
Compared different switch epochs:
2, 4, 8, 16, 32, 64, 128 (powers of 2)

After switching to the target distribution:

Best generalization achieved with a 2-stage curriculum

Spend first 128 epochs (half of total training time) on easier examples
Then switch to full target distribution

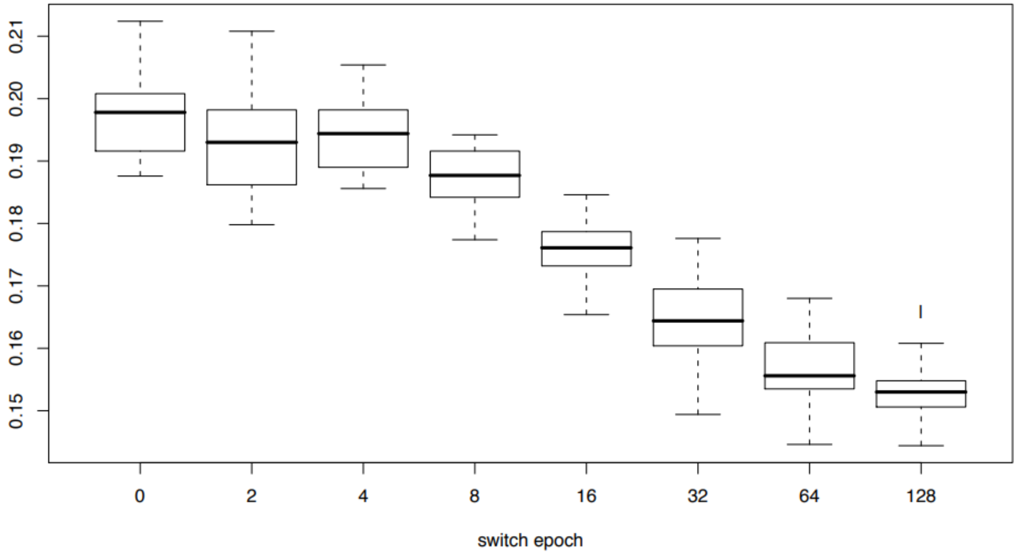
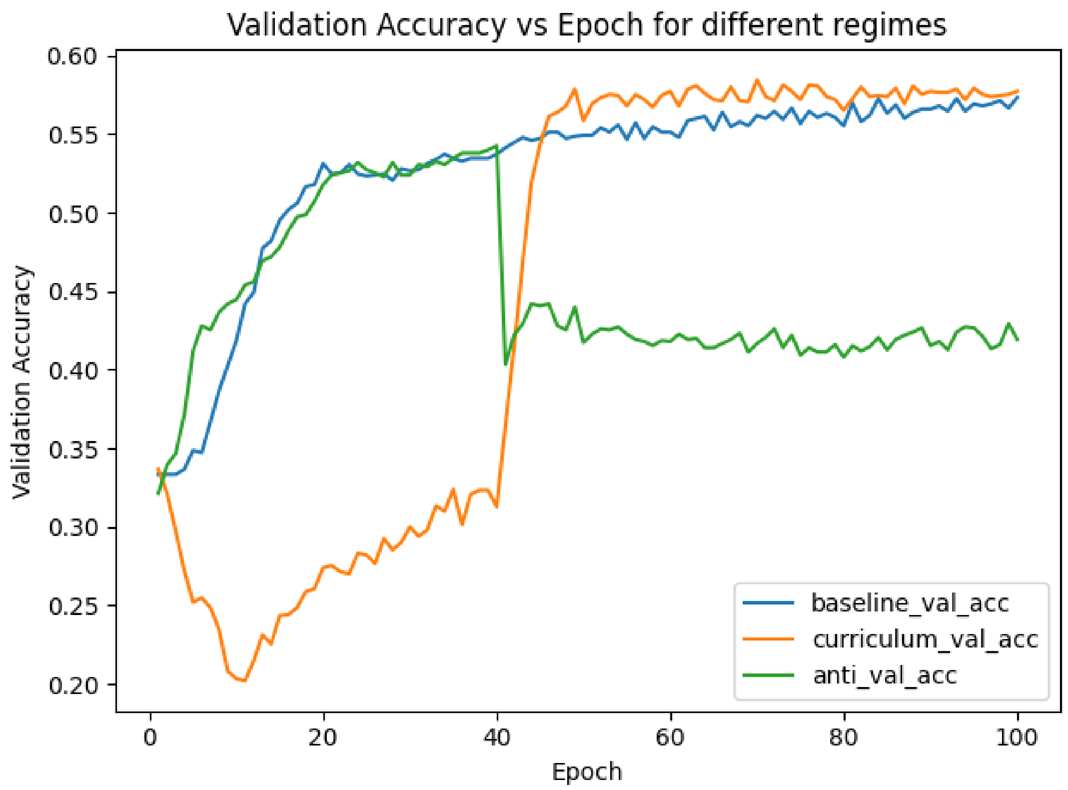


Figure : Box plot of test classification error distribution as a function of the "switch epoch"

Scan to check out the full implementation here

Click Link to check out the full implementation here



Plot : Validation accuracy comparison between Baseline, Curriculum Learning (CL), and Anti-Curriculum (Anti-CL) regimes.

How to apply this to your experiments and vision tasks ?

The real power of Curriculum Learning is:
It depends entirely on how well you design the scoring and pacing functions for your dataset and task.

The curriculum method,

- ❑ Depends on dataset
- ❑ Depends on model
- ❑ Depends on task
- ❑ Depends on data scarcity

How to apply Curriculum Learning ?

Curml: A curriculum machine learning library

Curriculum Learning can,

- ❑ Improve generalization
- ❑ Improve training stability
- ❑ Reduce training time
- ❑ Improve sample efficiency
- ❑ Be extremely useful in transfer learning
- ❑ Help when labeled data is limited

Scan to visit git repo : A collection of curriculum learning methods from literature.

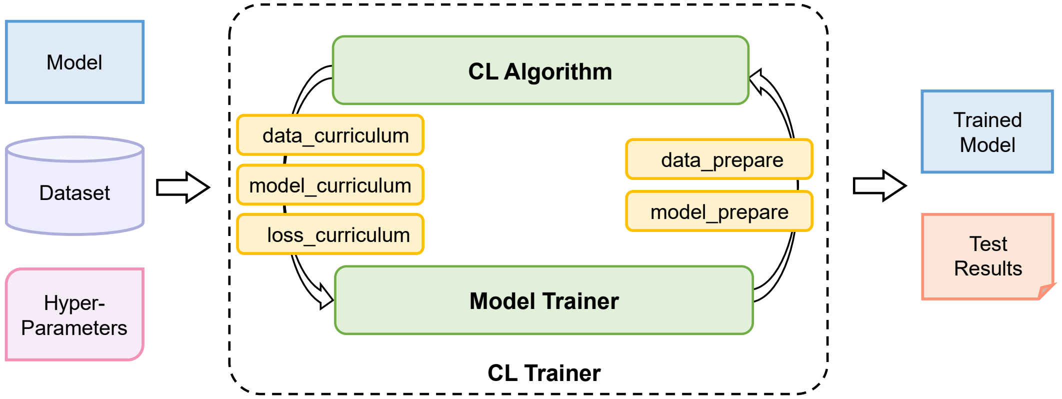


Figure : overall framework of the library

Predefined Scoring functions – Data Level and Model Level difficulty measures

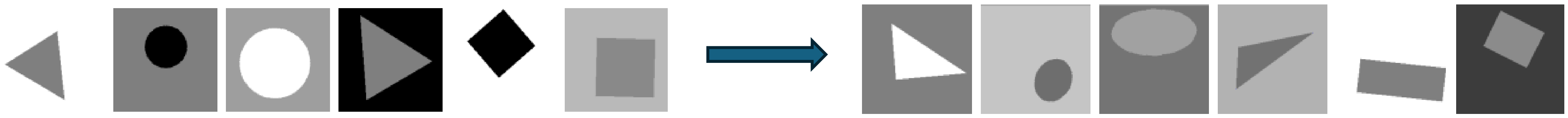
	scoring function/difficulty measure	**pacing function**
Shapes Experiment (3-class image classification)	Heuristic ranking of the dataset BasicShapes – “easy” GeomentricShapes – “hard”	Epochs 1–X: Easy data Epochs X–Y: Hard data
CBS	Difficulty is defined by **spatial frequency content** Low-frequency components → easy High-frequency components → hard Using a **global Gaussian low-pass filter**	A decay function σt=σ0 * α^t σ(t)↓0 over training

Predefined Scoring functions – Data Level and Model Level difficulty measures

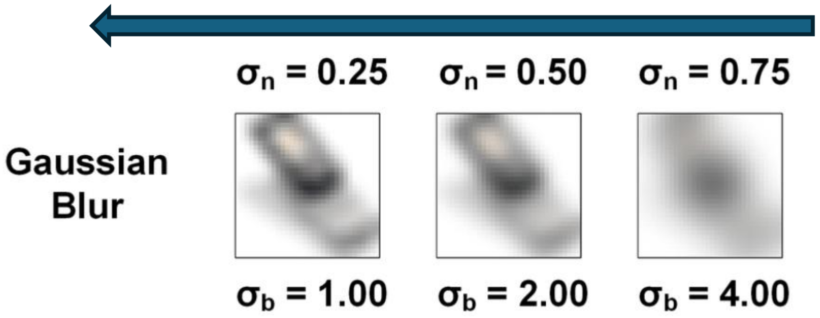
Shapes Experiment

Basic Shapes

Geometric shapes

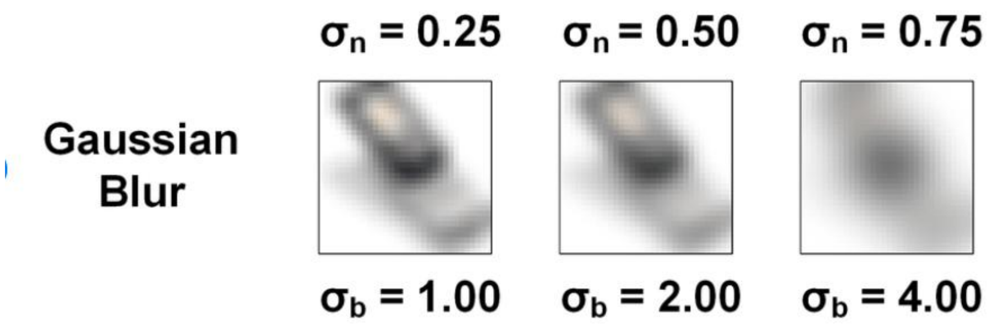


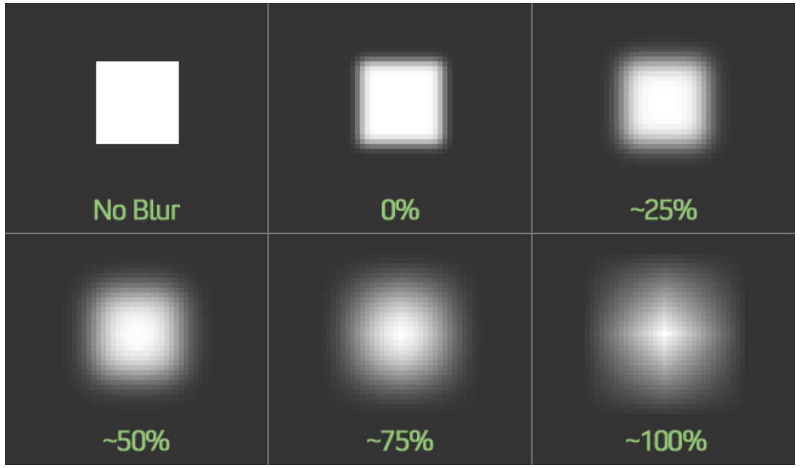
CBS



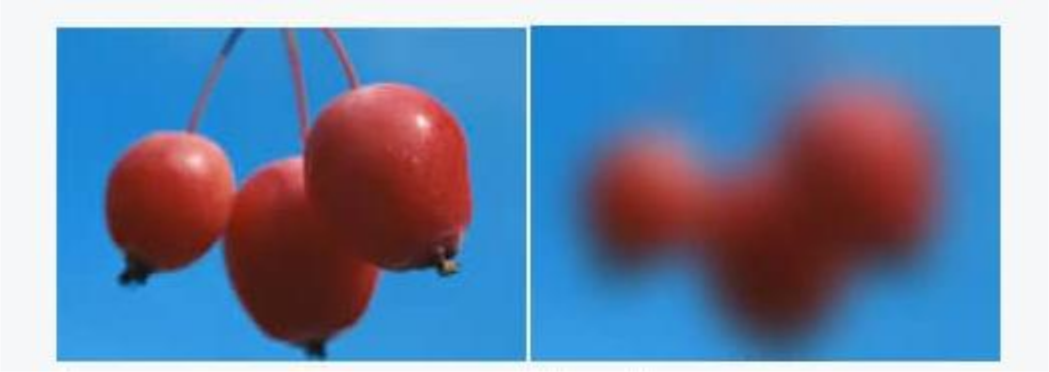
Curriculum By Smoothing (CBS)

[Sinha, Samarth, Animesh Garg, and Hugo Larochelle. *Advances in Neural Information Processing Systems* 33 (2020)]





High feature intensity → Low feature intensity



Why apply gaussian blur to images?

- ❑ Reduces high-frequency components
- ❑ Suppresses fine details and noise
- ❑ Preserves global structure

What is the difficulty measure here ?

- ❑ Amount of high-frequency information in an image

CBS ≠ Dataset augmentation or data-Level smoothing

Most curriculum learning methods decide which samples are easy or hard.

CBS defines difficulty as the amount of high-frequency information allowed in the feature maps.

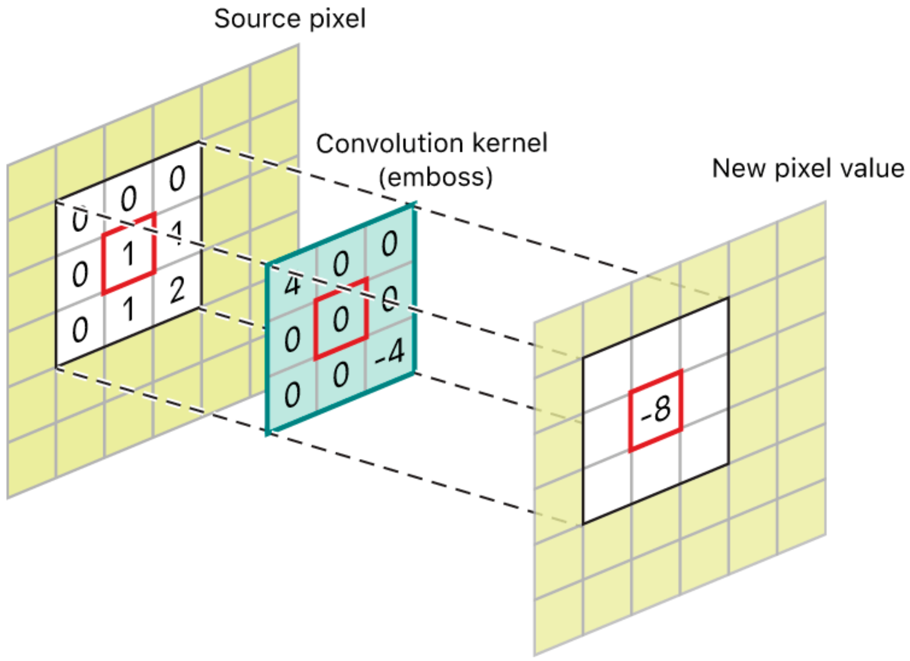


Figure : how gaussian blur is applied

How Does a Convolution Layer Work?

- ❑ The image is made of pixels (numbers)
- ❑ A small filter slides over the image
- ❑ It detects patterns (edges, textures)
- ❑ The output is a new image called a *feature map*

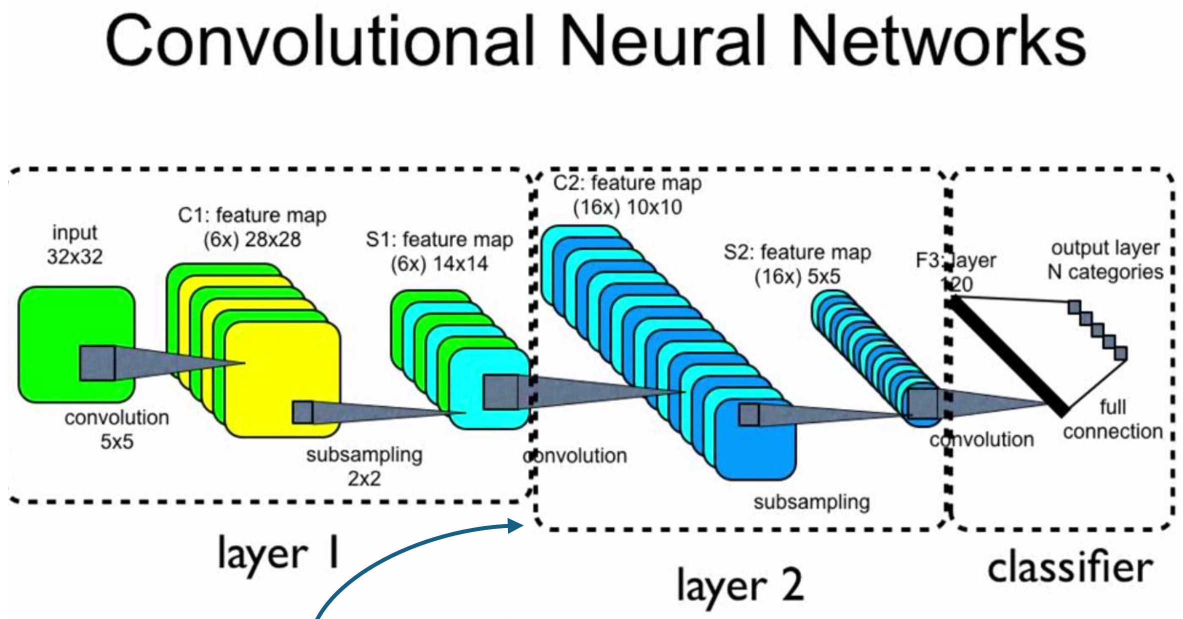


Image → Convolution → Feature Map → **Gaussian Blur** → BatchNorm → ReLU → Next Layer

Easy → Difficult

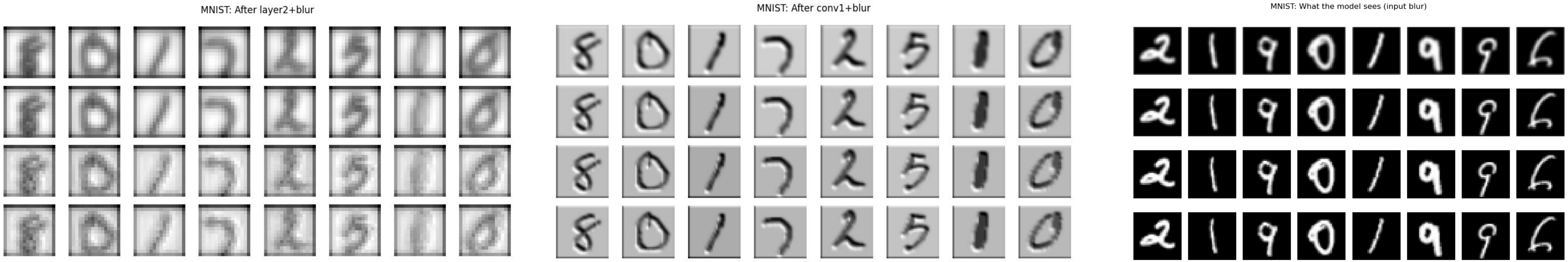


Figure : MNIST dataset

Predefined difficulty measure

High σ → strong blur → only low-frequency → *easy representations*
Low σ → weak blur → full HF detail → *hard representations*

Difficulty measure = amount of high-frequency detail allowed through.

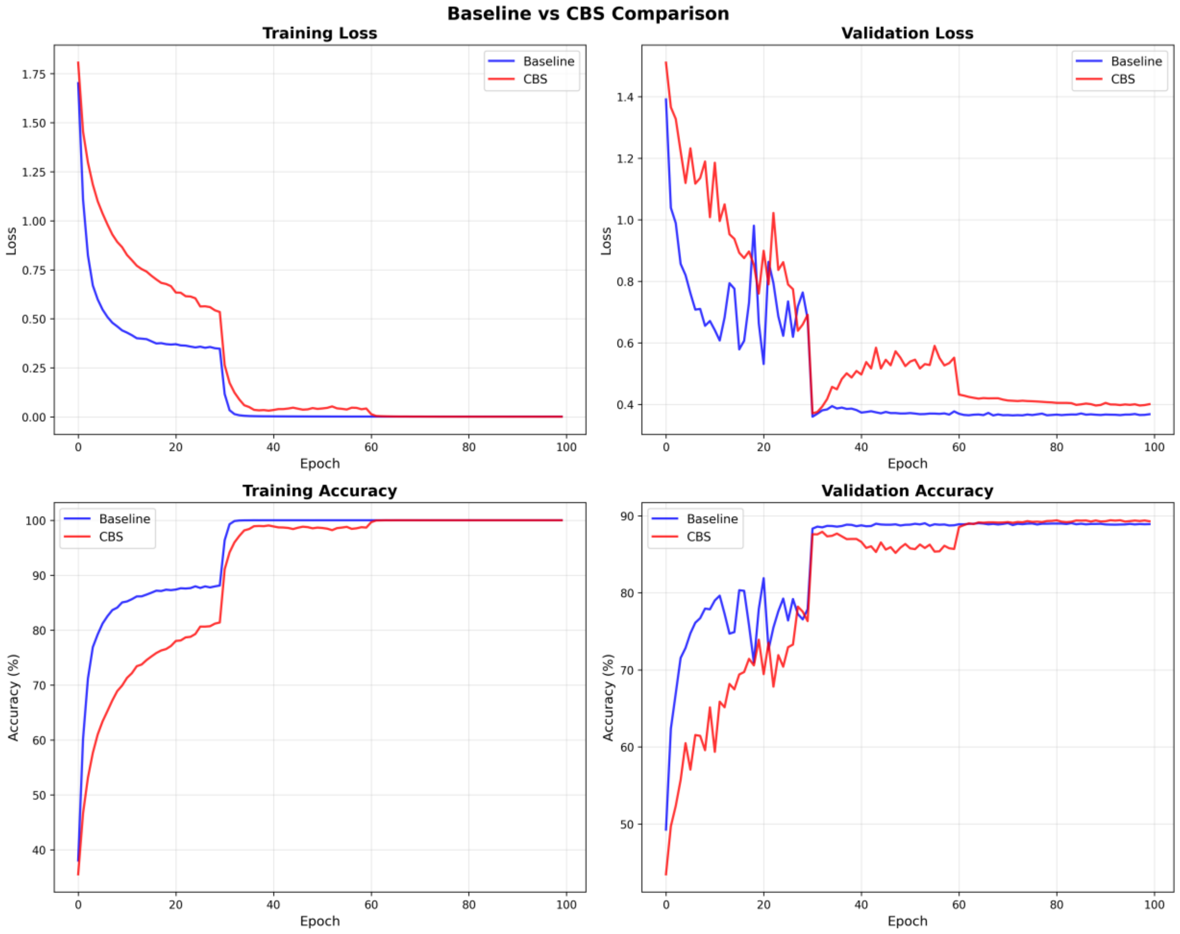
Predefined pacing function

σ(t) is **predefined**
σ(t) is **not influenced by the model**
σ(t) is **not adaptive**

Results from CBS

	SVHN	CIFAR10	CIFAR100
VGG-16	96.6 ± 0.2	85.8 ± 0.2	57.0 ± 0.2
VGG-16 + CBS	**97.0** ± 0.2	**88.9** ± 0.3	**61.4** ± 0.3
ResNet-18	97.2 ± 0.2	87.1 ± 0.3	62.4 ± 0.3
ResNet-18 + CBS	**98.7** ± 0.2	**90.2** ± 0.3	**65.4** ± 0.2
Wide-ResNet-50	97.7 ± 0.1	91.8 ± 0.1	73.3 ± 0.1
Wide-ResNet-50 + CBS	**98.3** ± 0.3	**93.9** ± 0.1	**75.9** ± 0.2
ResNeXt-50	97.7 ± 0.2	93.1 ± 0.1	74.1 ± 0.3
ResNeXt-50 + CBS	**99.0** ± 0.2	**95.1** ± 0.2	**77.0** ± 0.1

Figure : Results table from CBS



Plot : Baseline vs Curriculum with CBS