IMCOM 2025

International Conference on Ubiquitous Information Management and Communication

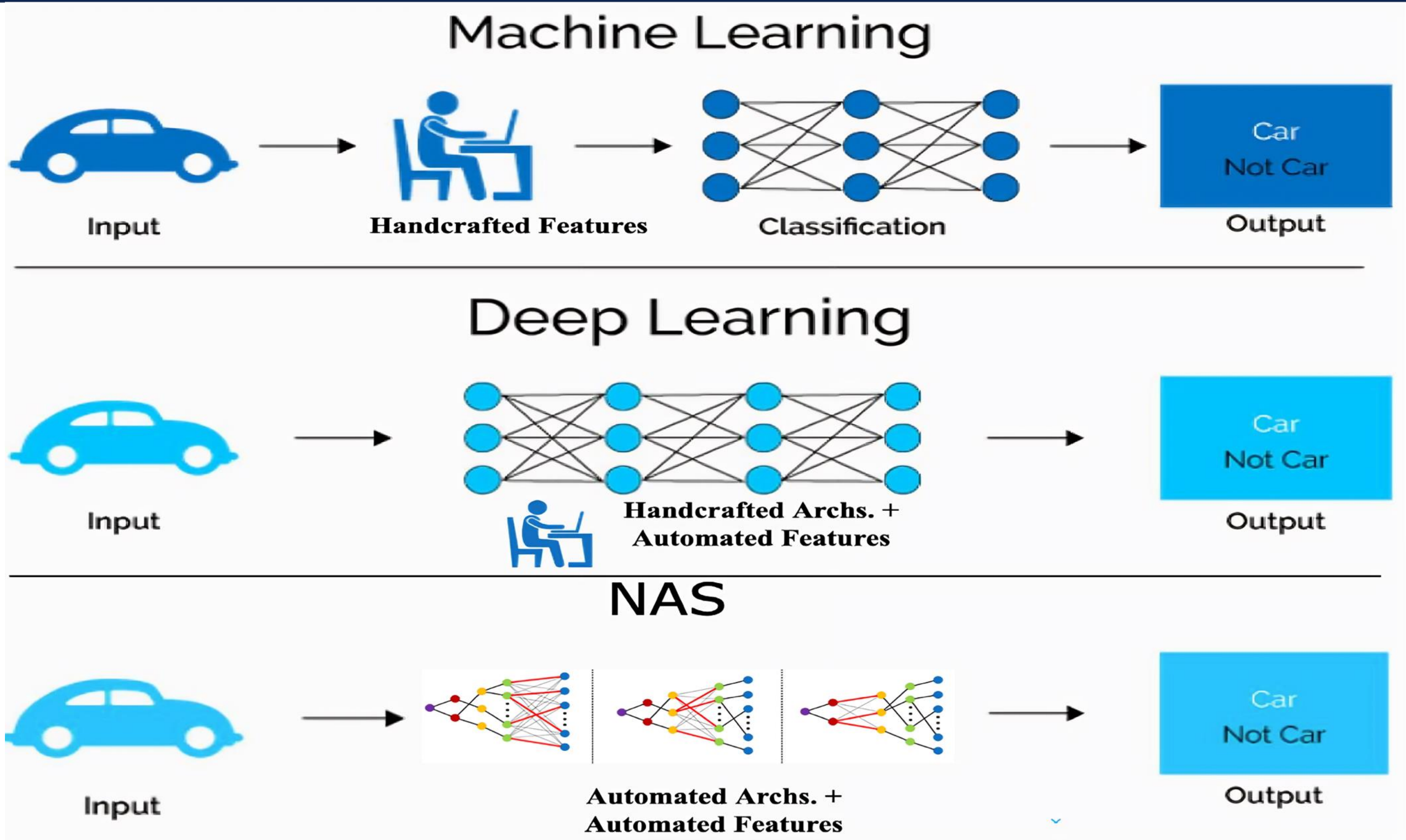
January 3~5, 2025 – Bangkok, Thailand

Efficient Self-Supervised Neural Architecture Search

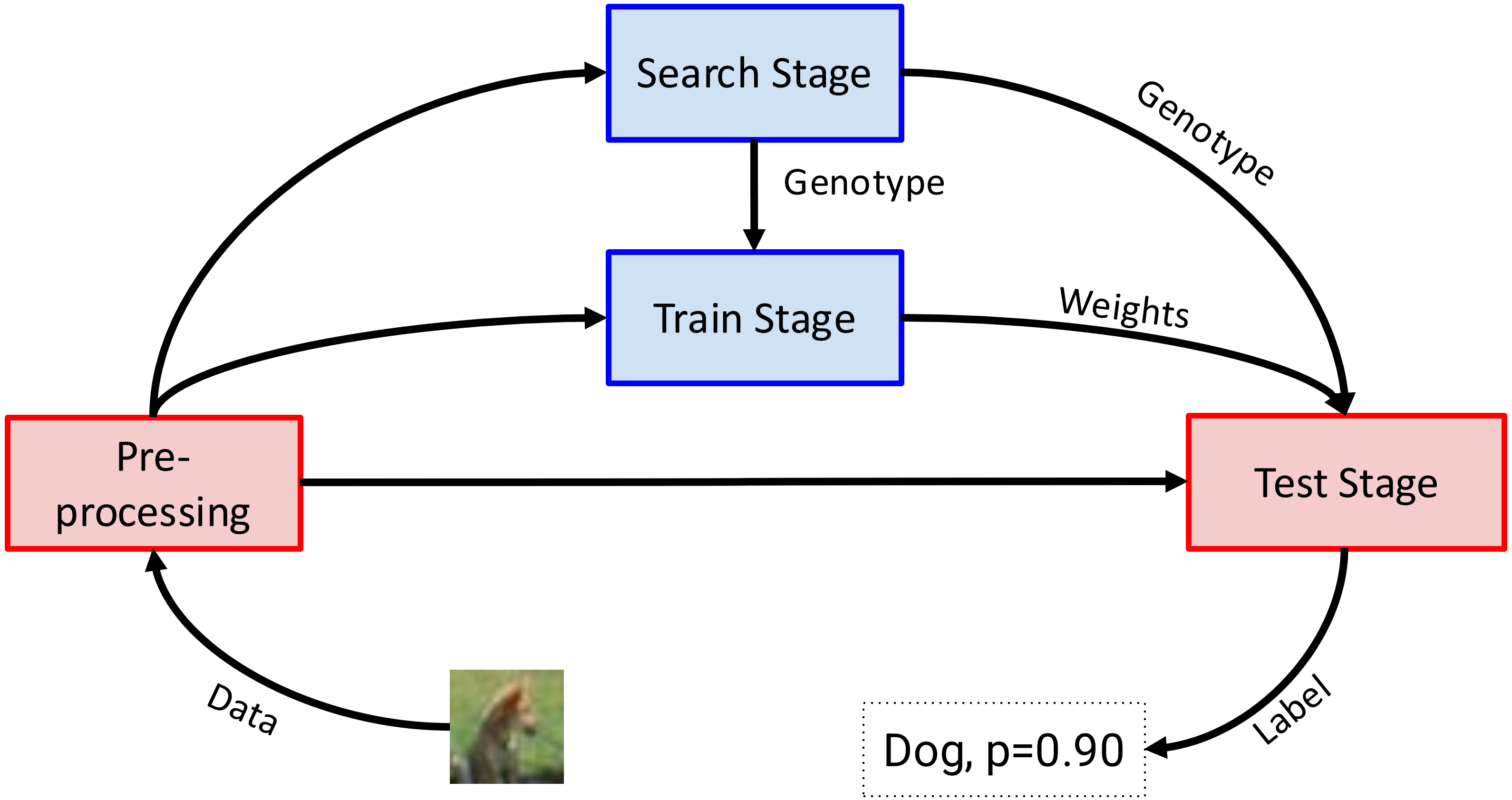
Presenter: Sri Aditya Deevi

ISBN: 979-8-3315-0782-4

Motivation


Machine Learning
Input
Handcrafted Features
Classification
Car
Not Car
Output
Deep Learning
Input
Handcrafted Archs. +
Automated Features
Car
Not Car
Output
NAS
Input
Automated Archs. +
Automated Features
Car
Not Car
Output

Bird's Eye View



Proposed Approach : Key Elements

Differentiable Architecture Search

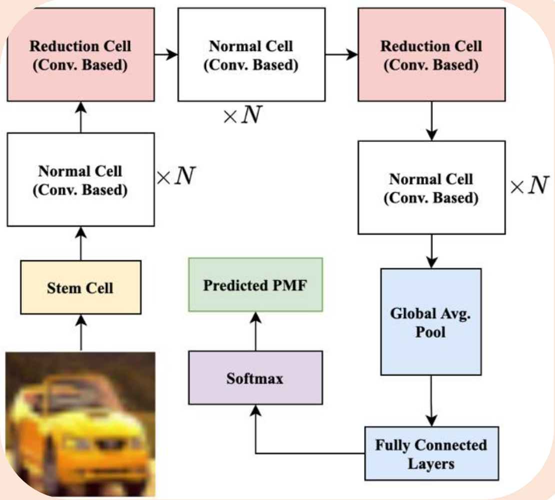
- Search of Neural Architectures based Bilevel optimization using gradients.

$$\min_\alpha \quad \mathcal{L}_{val}\left(w^*(\alpha), \alpha\right)$$
$$\text{s.t.} \quad w^*(\alpha) = \text{argmin}_w \ \mathcal{L}_{train}(w, \alpha)$$

$w \rightarrow$ Network Weights

$\alpha \rightarrow$ Architecture (Ops) Weights

- Micro search for convolutional operation based computational “cells”.

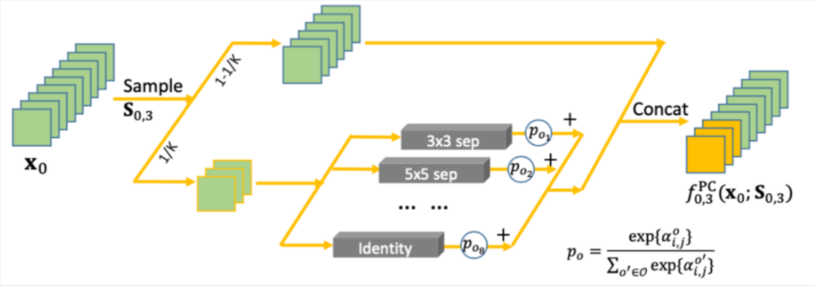


Partial Channel Connections

- Idea: Randomly sample (1/K) of the total channels for operation selection.

$$f_{i,j}^{PC}\left(\mathbf{x}_i; \mathbf{S}_{i,j}\right) = \sum_{o \in \mathcal{O}} \frac{\exp\left\{\alpha_{i,j}^{o}\right\}}{\sum_{o' \in \mathcal{O}} \exp\left\{\alpha_{i,j}^{o'}\right\}} \cdot o\left(\mathbf{S}_{i,j} * \mathbf{x}_i\right) + \left(1 - \mathbf{S}_{i,j}\right) * \mathbf{x}_i \quad \Big| \quad \mathbf{S}_{i,j} \rightarrow \text{Sampling Mask}$$

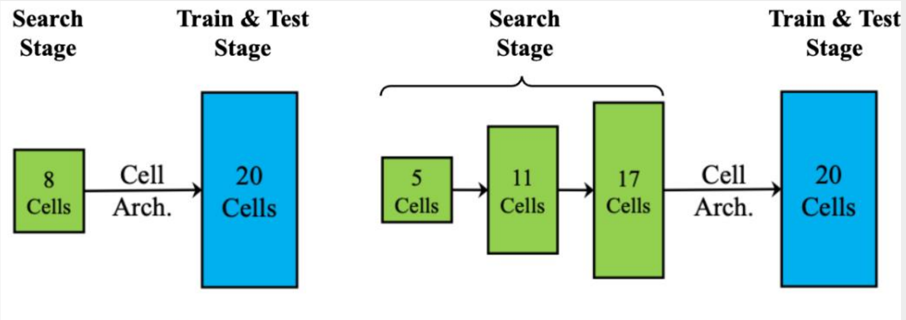
Here, K = 4



Makes the search efficient!

Progressive Search

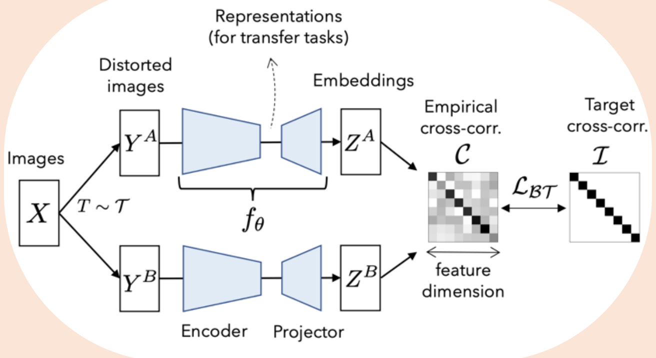
- Bridges the “Depth Gap” between search and evaluation (Certain ops. might be preferred in deep networks)



(a) DARTS (b) Progressive DARTS

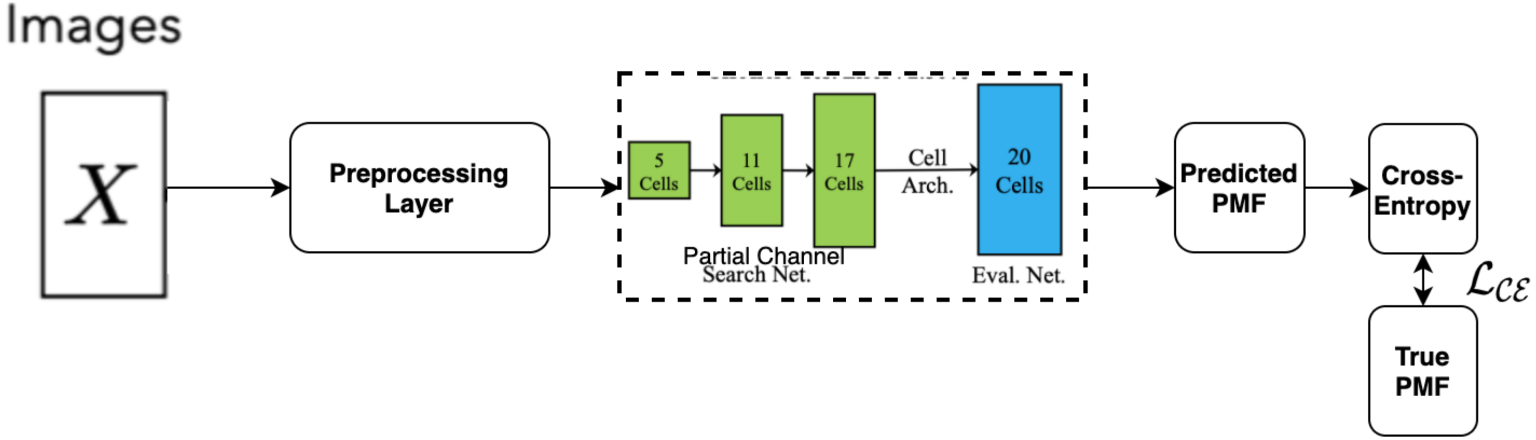
Barlow Twins Self-Supervision

- Optimizing loss to make cross-correlation close to identity matrix:
  - ✓ DNN learns latent feature vectors
  - ✓ Ignore/generalize distortions
  - ✓ Reduces redundancy



$$\mathcal{L}_{\mathcal{BT}} \triangleq \underbrace{\sum_i \left(1 - \mathcal{C}_{ii}\right)^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} \mathcal{C}_{ij}^2}_{\text{redundancy reduction term}} \quad \Big| \quad \mathcal{C}_{ij} \triangleq \frac{\sum_b z_{b,i}^A z_{b,j}^B}{\sqrt{\sum_b \left(z_{b,i}^A\right)^2} \sqrt{\sum_b \left(z_{b,j}^B\right)^2}}$$

Proposed Approach : Illustrative Overview



Vanilla PCP DARTS

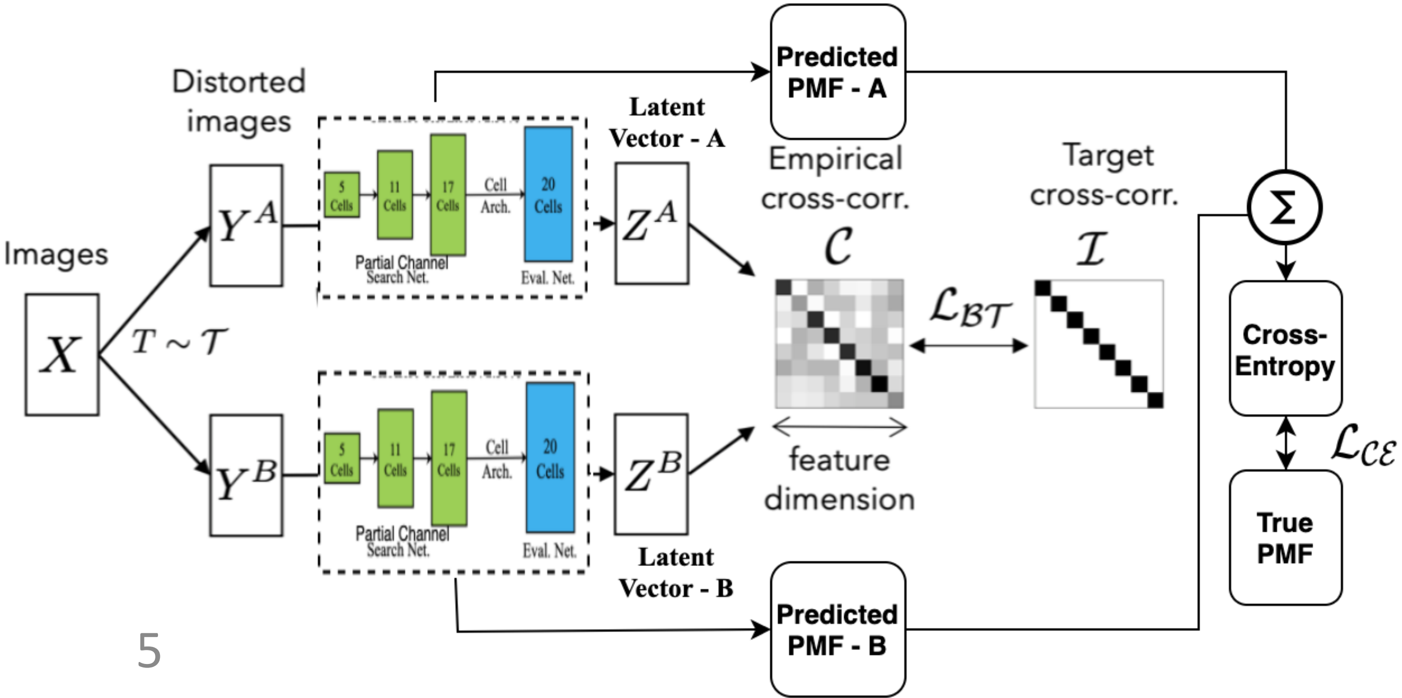
- Salient Features are:
  - ✓ DARTS bilevel optimization framework
  - ✓ Partial Channel connections for efficiency
  - ✓ Progressive search for bridging depth gap
  - ✓ Supervised Loss for Search and Classification

Self-Supervised PCP DARTS

- Salient Features are:
  - ✓ DARTS bilevel optimization framework
  - ✓ Partial Channel connections for efficiency
  - ✓ Progressive search for bridging depth gap
  - ✓ Mix. of Self-Supervised & Supervised Loss for Search:

$$\mathcal{L} = \text{ss\_factor} \times \mathcal{L}_{\mathcal{BT}} + (1 - \text{ss\_factor})\mathcal{L}_{\mathcal{CE}}$$

  - ✓ Supervised Loss for Classification

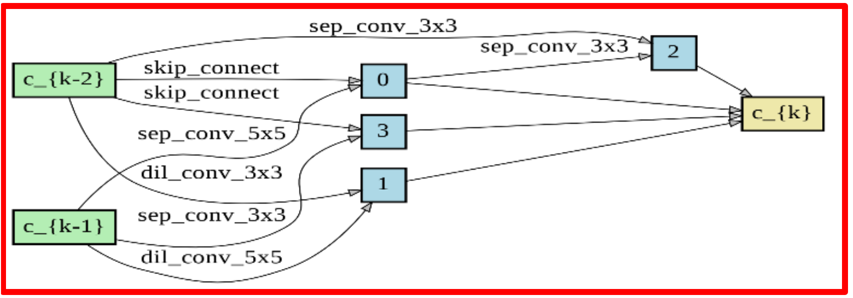


Notable Results & Analyses

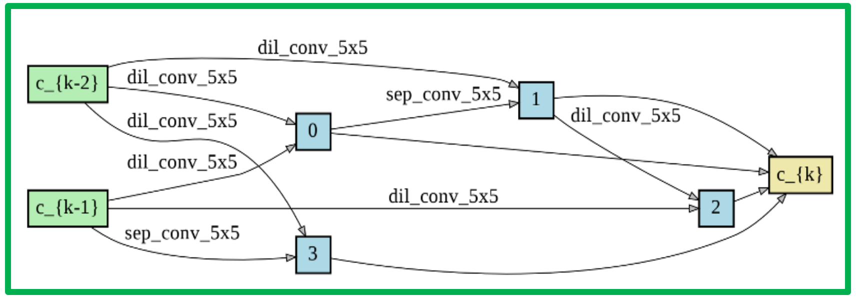
- Based on some experiments conducted on CIFAR-10 dataset we found that (quantitatively) :
  - Time taken for Architecture Search: PC-DARTS ≲ PCP-DARTS < P-DARTS
  - Test Accuracy : PC-DARTS < PCP-DARTS ≲ P-DARTS

PC → Partial Channel connections
P → Progressive

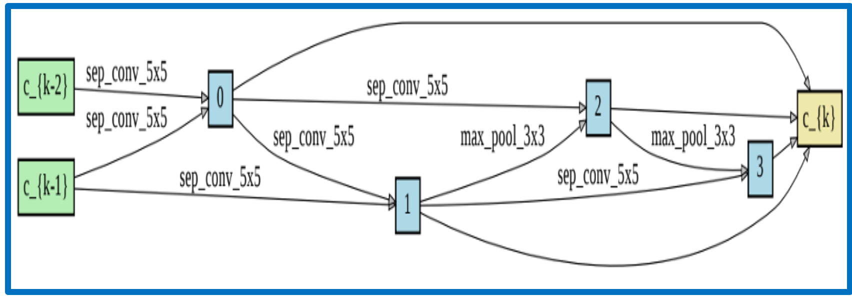
Vanilla PC-DARTS Normal Cell



Vanilla PCP-DARTS Normal Cell



SS(0.75) PC-DARTS Normal Cell



- Some interesting observations comparing the above searched architectures are:
  - Vanilla PCP-DARTS has less number of skip connections and more depth than Vanilla PC-DARTS.
  - Self-Supervised Architecture Search leads the network to have more depth.
  - Parameter less operations like skip-connections are preferred in PC-DARTS as compared to PCP-DARTS.

Notable Results & Analyses

- We also tested the robustness of our methods by generating adversarial test examples:

Fast Gradient Sign Method (FGSM)

$$X^{adv} = X + \epsilon \operatorname{sign}\left(\nabla_X J\left(X, Y_{\text{true}}\right)\right)$$

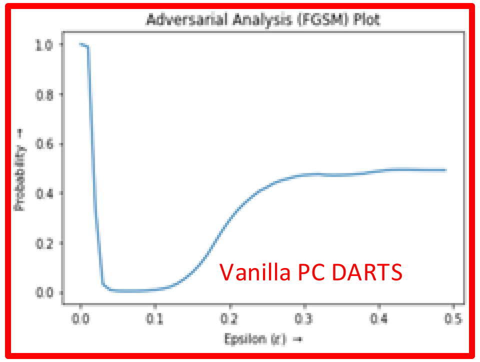
where

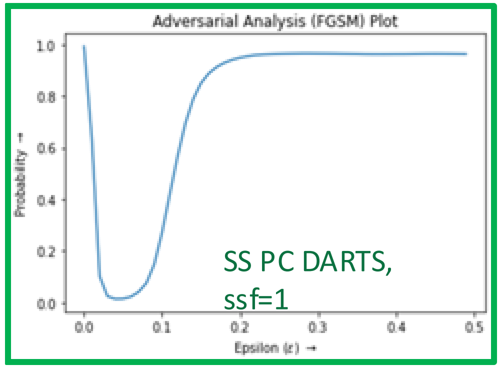
$X =$ original (clean) input

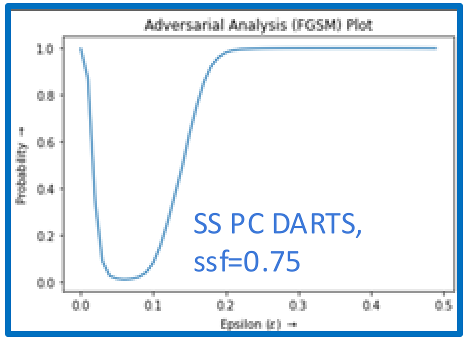
$X_{adv} =$ adversarial input

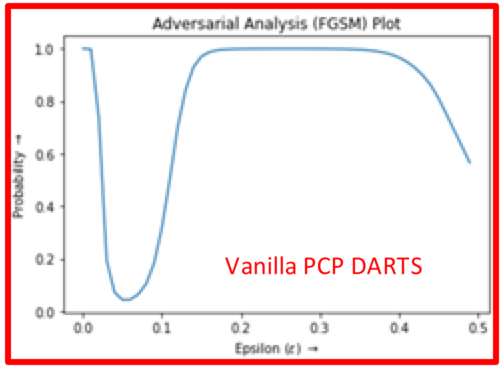
$\epsilon =$ magnitude of adversarial perturbation

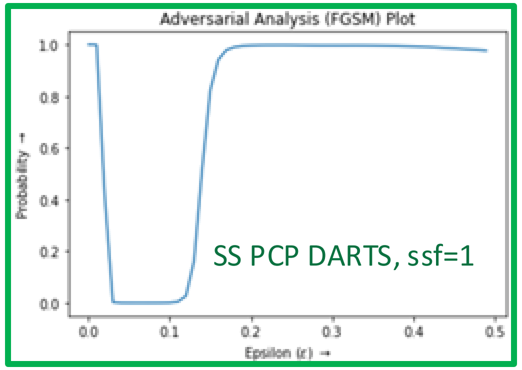
$\nabla_X J\left(X, Y_{true}\right) =$ gradient of loss function w.r.t to input (X)

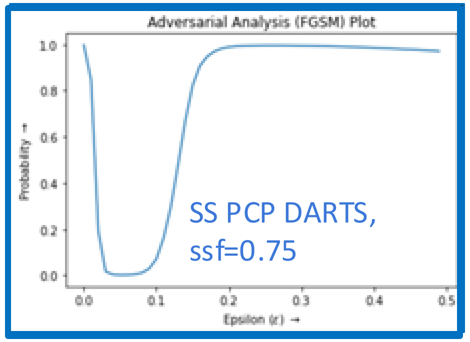












Conclusion

- We presented efficient neural architecture search algorithms to address the high resource demands of traditional handcrafted neural architectures.

- We conducted experiments in both fully supervised and self-supervised settings, utilizing a combined loss function of supervised cross-entropy and self-supervision loss to guide the search for optimal architectures.

- We analysed performance on CIFAR-10, demonstrating that the proposed methodology balances time and accuracy, achieving results with less than 3% test error, close to state-of-the-art benchmarks.

- We provided interesting analyses that indicate the effectiveness of our proposed methods while transfer-training as well as their robustness in presence of adversarial noise.

THANK YOU!

Efficient Self-Supervised Neural Architecture Search

Sri Aditya Deevi*
Mission Simulation Group
U.R. Rao Satellite Centre, ISRO
Bengaluru, India
saditya@ursc.gov.in

Asish Kumar Mishra*
Mission Simulation Group
U.R. Rao Satellite Centre, ISRO
Bengaluru, India
asishkm@ursc.gov.in

Deepak Mishra
Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram, India
deepak.mishra@iist.ac.in

Ravi Kumar L
Mission Simulation Group
U.R. Rao Satellite Centre, ISRO
Bengaluru, India
rkkumarl@ursc.gov.in

Bharat Kumar G V P
Mission Simulation Group
U.R. Rao Satellite Centre, ISRO
Bengaluru, India
bharat@ursc.gov.in

Murali Krishna Bhagavan G
Mission Simulation Group
U.R. Rao Satellite Centre, ISRO
Bengaluru, India
bhagavan@ursc.gov.in

dsriaditya999@gmail.com
https://www.sriadityadeevi.com