Report

Expert Lecture

(Session: - 2022-23)

Topic of Talk: Hardware-aware Neural Architecture Search (HW-NAS)

Keynote speaker - Prof. (Dr.) Arun K. Somani, Associate Dean for Research, College of Engineering, Iowa State University, Ames, USA.

Date: - 04/02/2022

Keynote Session Time: 09:00 am - 09:30 am

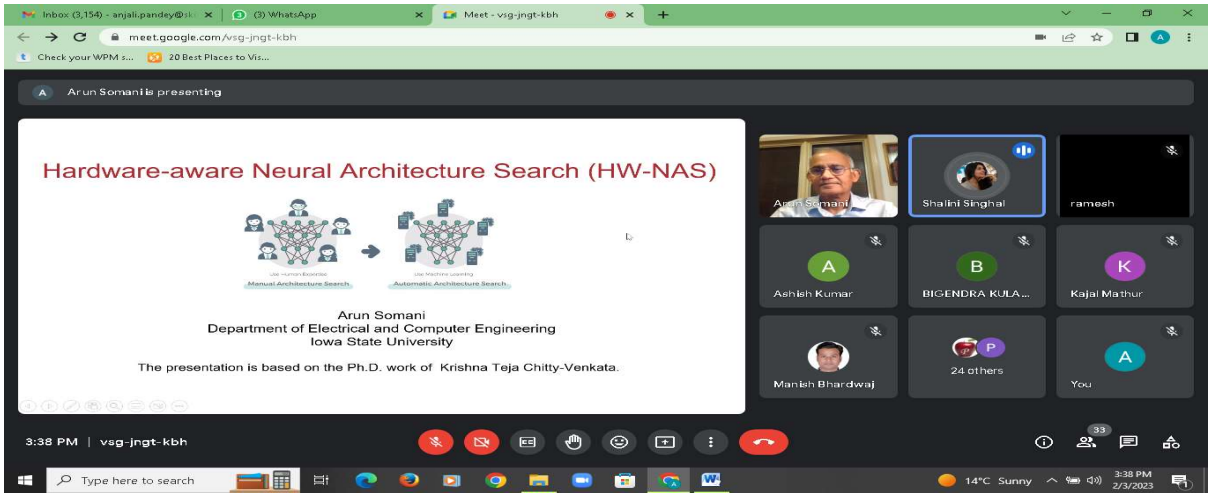
Link of the Session: - https://meet.google.com/vsg-jngt-kbh

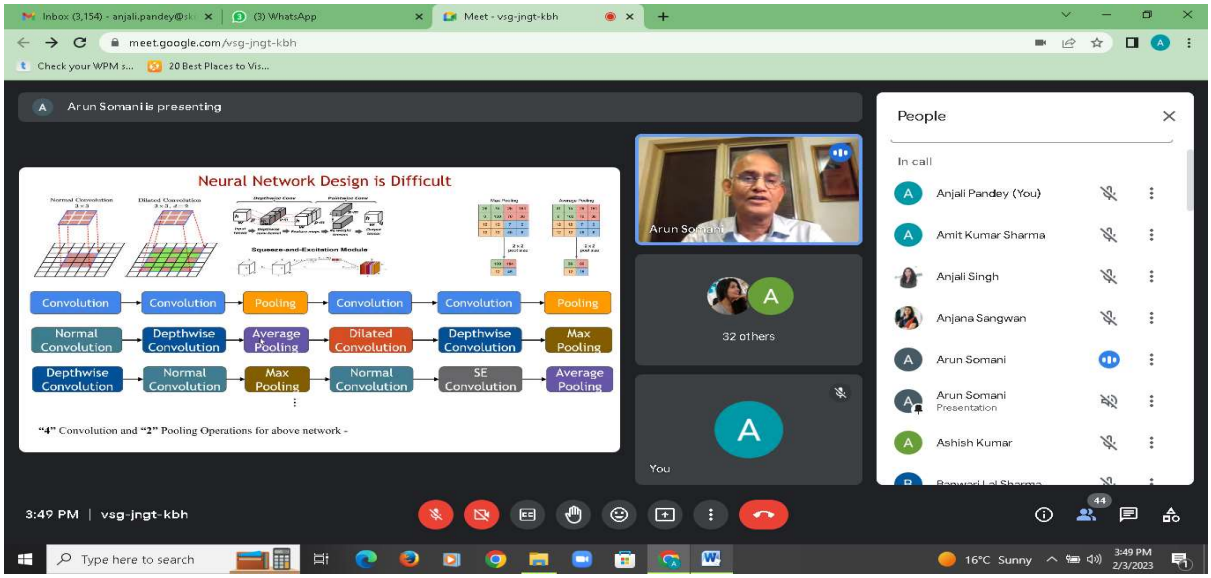
No. of Participant:-103

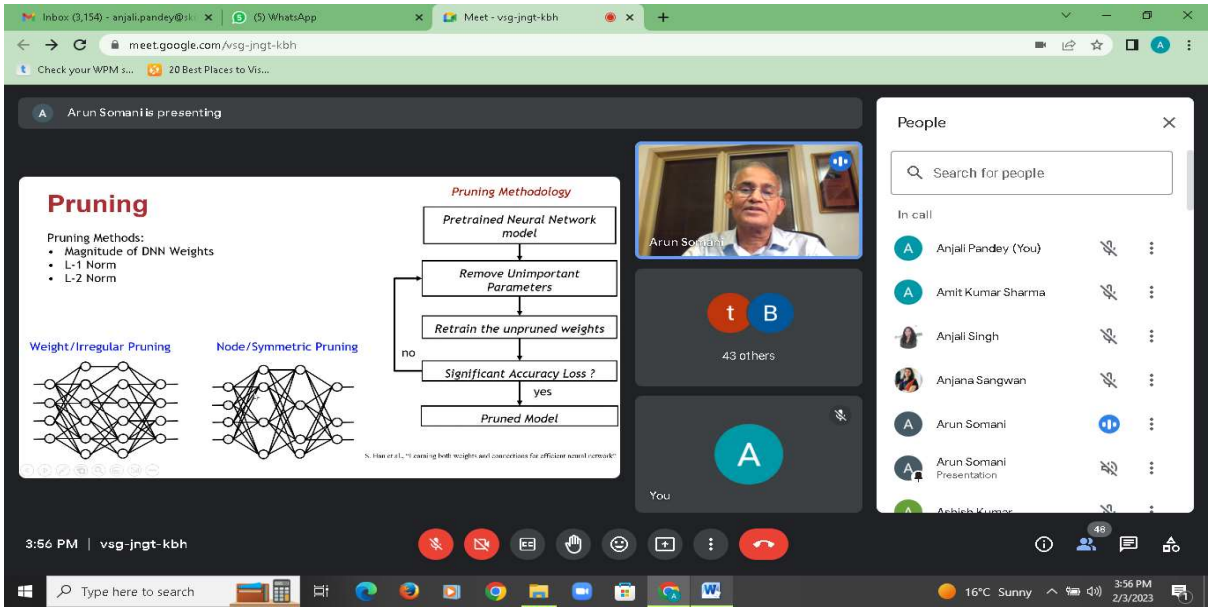
Summary of the keynote session:-

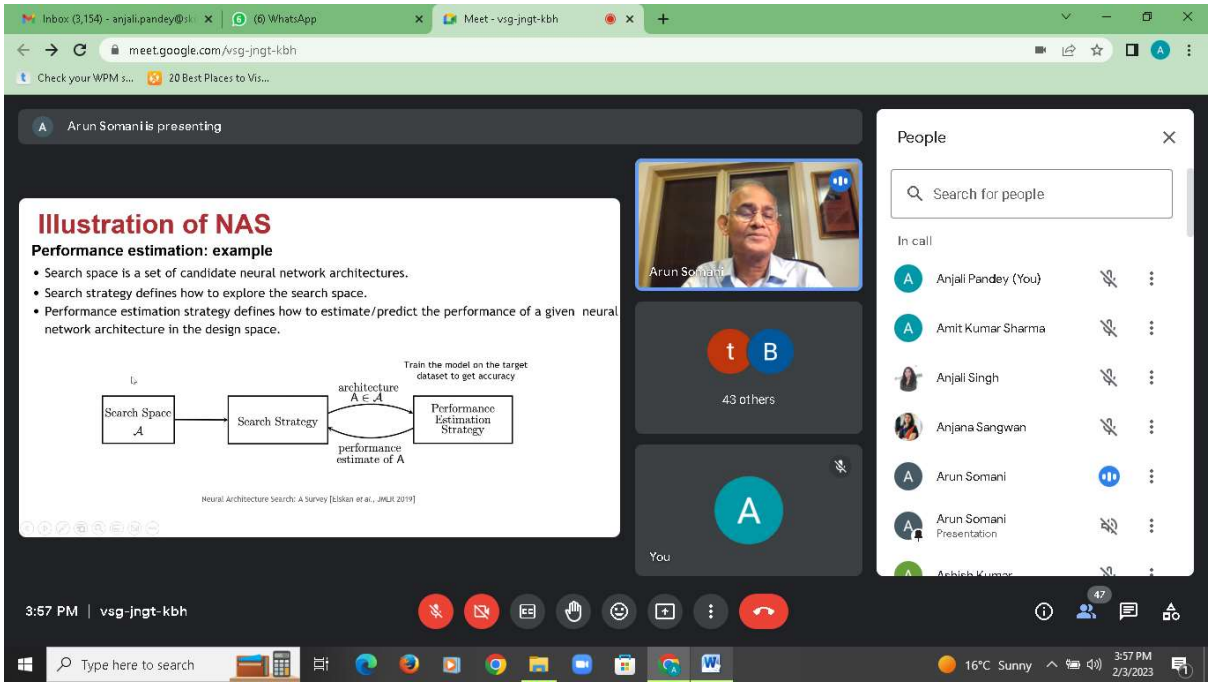
- Prof Somani, has talked about the Hardware-aware Neural Architecture Search.
- He has talked about convolutions operations, manually designed neural networks and designing of neural networks.
- He has talked about pruning & its methods, quantization.
- He has talked about hardwares can be used for deep learning.
- He has emphasized on the neural architecture search – Reinforcement based, evolutionary, DARTS.

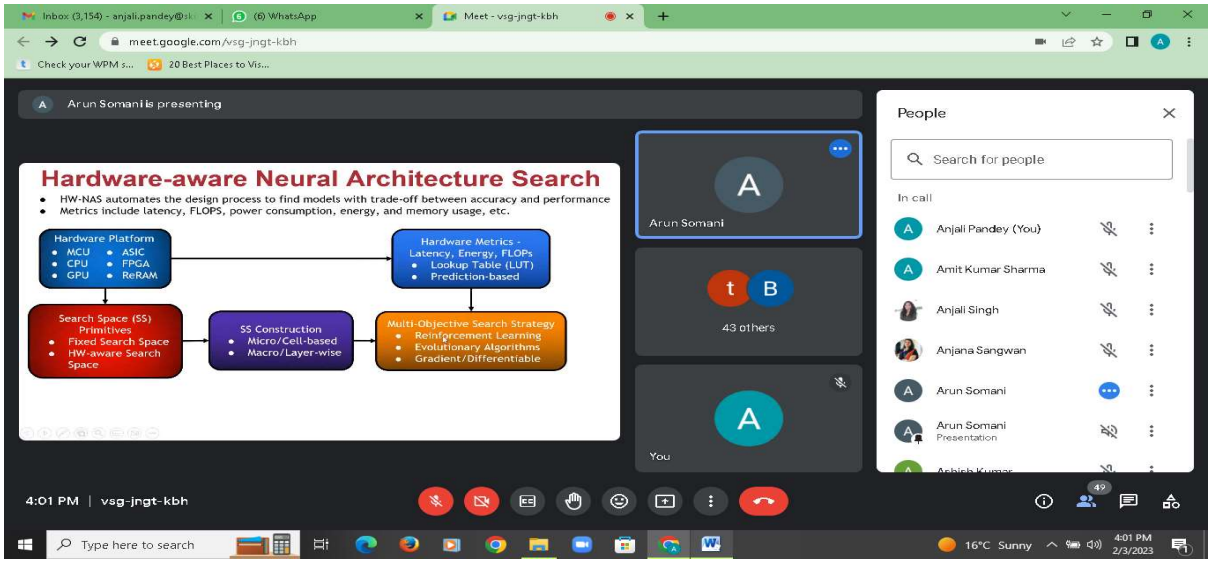
Screen Shot of the Session:-

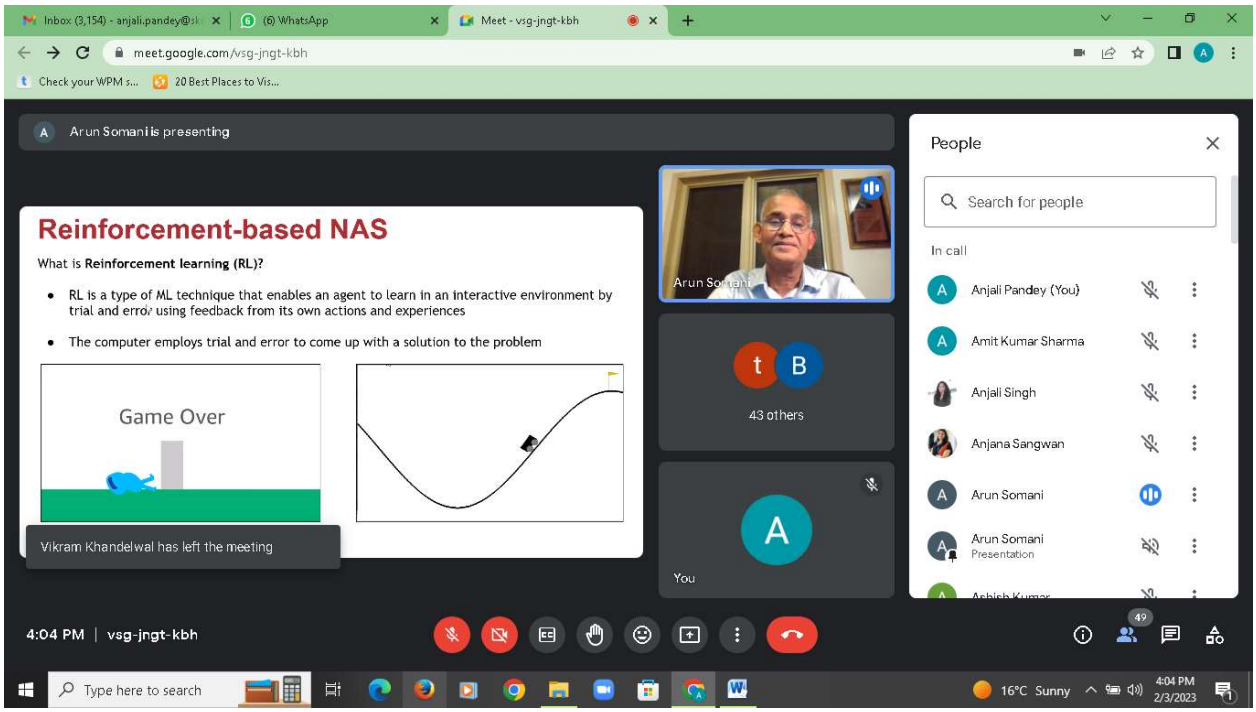





Arun Somani is presenting
Pruning
Pruning Methods:
• Magnitude of DNN Weights
• L-1 Norm
• L-2 Norm
Pruning Methodology
Pretrained Neural Network model
Remove Unimportant Parameters
Retrain the unpruned weights
Significant Accuracy Loss ?
no
yes
Pruned Model
Weight/Irregular Pruning
Node/Symmetric Pruning
Arun Somani
43 others
You
People
Search for people
In call
Anjali Pandey (You)
Amit Kumar Sharma
Anjali Singh
Anjana Sangwan
Arun Somani
Arun Somani Presentation
3:56 PM | vsg-jngt-kbh


Arun Somani is presenting
Illustration of NAS
Performance estimation: example
• Search space is a set of candidate neural network architectures.
• Search strategy defines how to explore the search space.
• Performance estimation strategy defines how to estimate/predict the performance of a given neural network architecture in the design space.
Search Space A
Search Strategy
Performance Estimation Strategy
architecture A ∈ 𝒜
performance estimate of A
Train the model on the target dataset to get accuracy
Arun Somani
43 others
You
People
Search for people
In call
Anjali Pandey (You)
Amit Kumar Sharma
Anjali Singh
Anjana Sangwan
Arun Somani
Arun Somani Presentation
3:57 PM | vsg-jngt-kbh


Arun Somani is presenting
Hardware-aware Neural Architecture Search
• HW-NAS automates the design process to find models with trade-off between accuracy and performance
• Metrics include latency, FLOPS, power consumption, energy, and memory usage, etc.
Hardware Platform
MCU ASIC
CPU FPGA
GPU ReRAM
Hardware Metrics - Latency, Energy, FLOPs
Lookup Table (LUT)
Prediction-based
Search Space (SS) Primitives
Fixed Search Space
HW-aware Search Space
SS Construction
Micro/Cell-based
Macro/Layer-wise
Multi-Objective Search Strategy
Reinforcement Learning
Evolutionary Algorithms
Gradient/Differentiable
Arun Somani
43 others
You
People
Search for people
In call
Anjali Pandey (You)
Amit Kumar Sharma
Anjali Singh
Anjana Sangwan
Arun Somani
Arun Somani Presentation
4:01 PM | vsg-jngt-kbh



GROUP PHOTOGRAPH