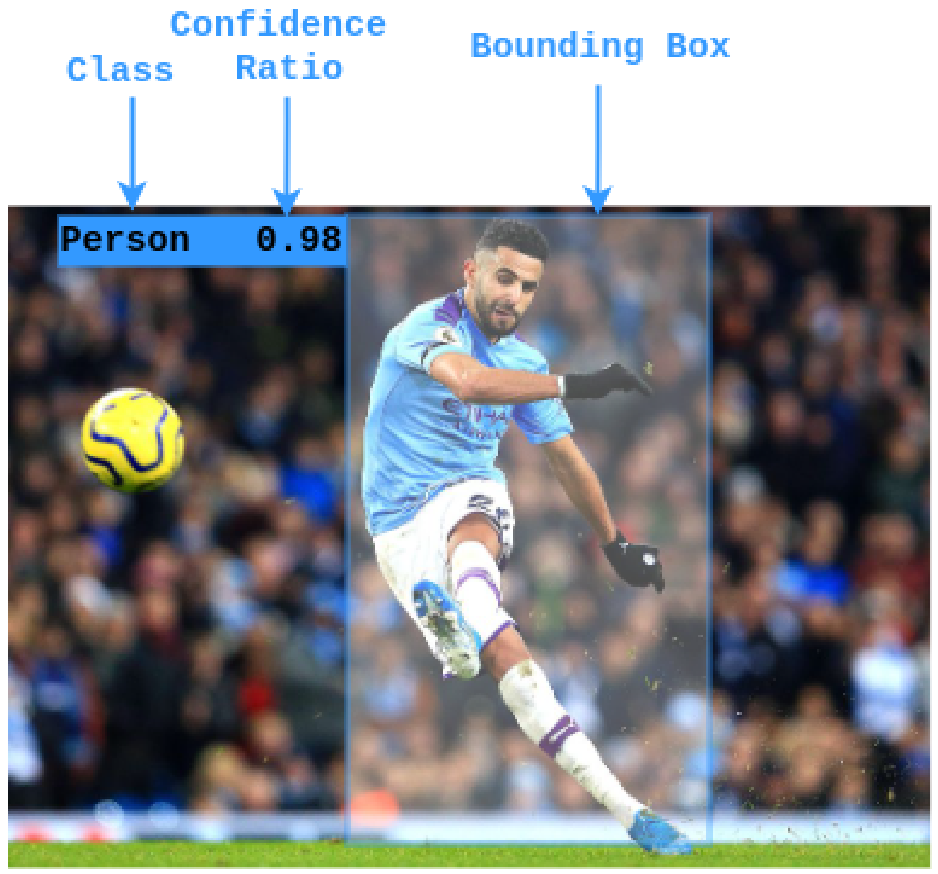
HyT-NAS: Hybrid Transformers Neural Architecture Search for Edge Devices

Date: 13/10/2022

MECHARBAT Lotfi Abdelkrim, BENMEZIANE Hadjer,
OUARNOUGHI Hamza, NIAR Smail

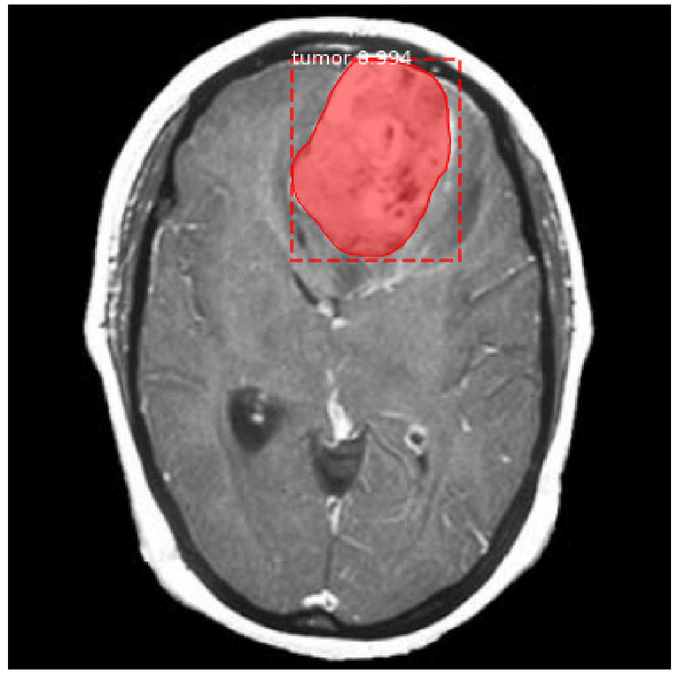
Introduction

Visual Object Recognition



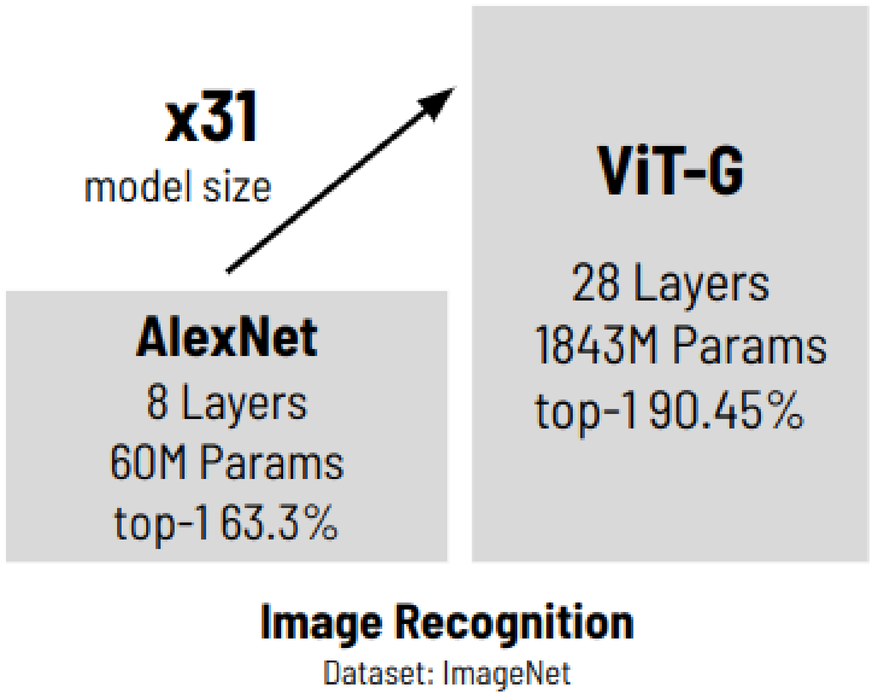
Identification + Localization

Applications



Deep Learning is the dominant approach for visual object recognition

Characteristics of deep learning models

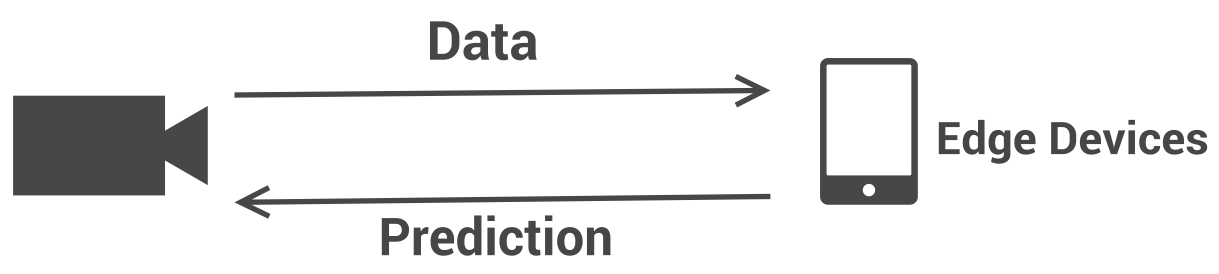


(Benmeziane et al, 2021)

- High accuracy in various fields, including object recognition.
- Extremely flexible due to the wide variety of hyperparameters that control them.
- High computational and memory complexity.

Motivation

Edge AI



- Unreliable (depends on network quality).
- Slow process for real-time applications.
- Not suitable for critical applications.

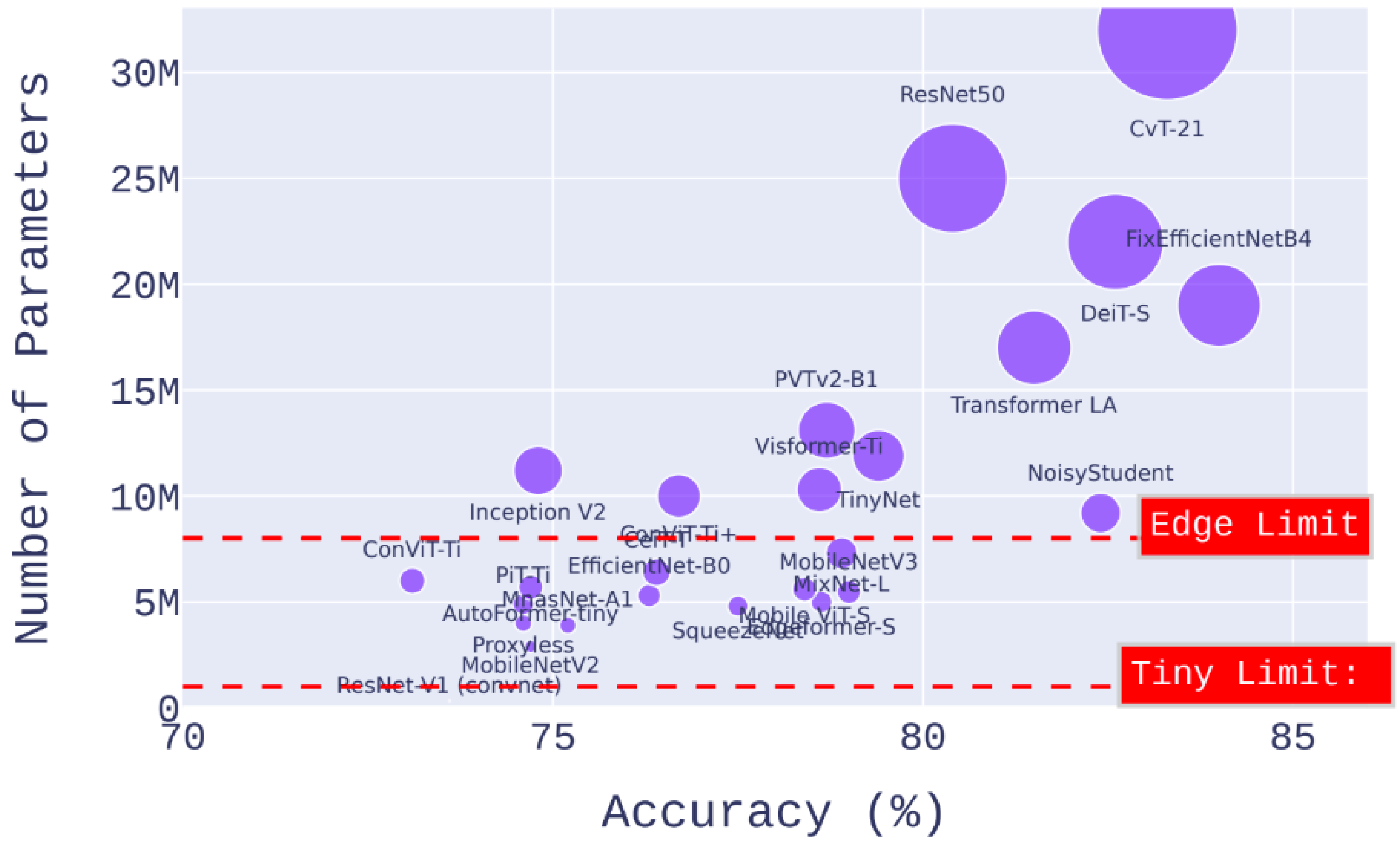
- More reliable.
- No data transfer over the network.
- Preserve confidentiality.

⬢ Challenges of edge AI ⬢

Gigantic architectures, models are too big to fit in Edge devices.

Huge computational complexity, not fast enough for inference in Edge

High power consumption, drains the limited power source (battery) of Edge devices.

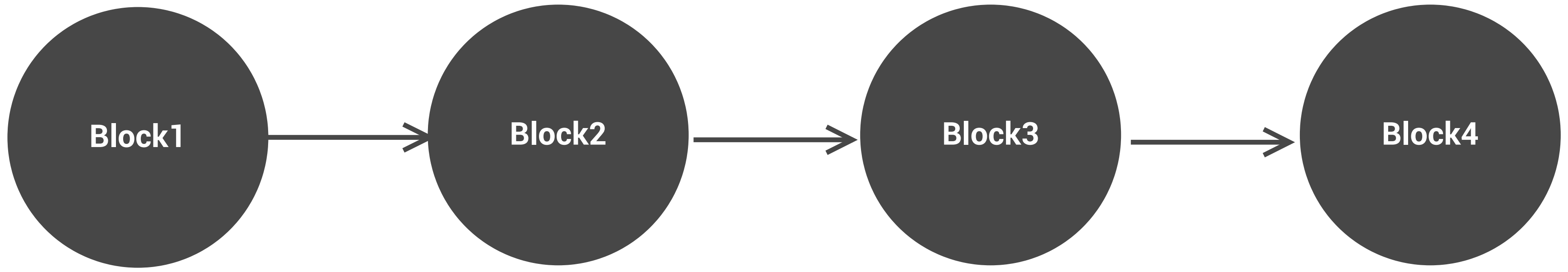


Motivation

Hyperparameters optimization

- Why is it difficult to make a good choice of hyperparameters manually?

The space of possible configurations is of immense size



Typical architecture

If Each Block has 5 Hyperparameters to set, each with 4 possible values

Then Size of the space of possible configurations= 1.099 * 10^12

With 1s/eval, the exploration of this space requires more than 30,000 years

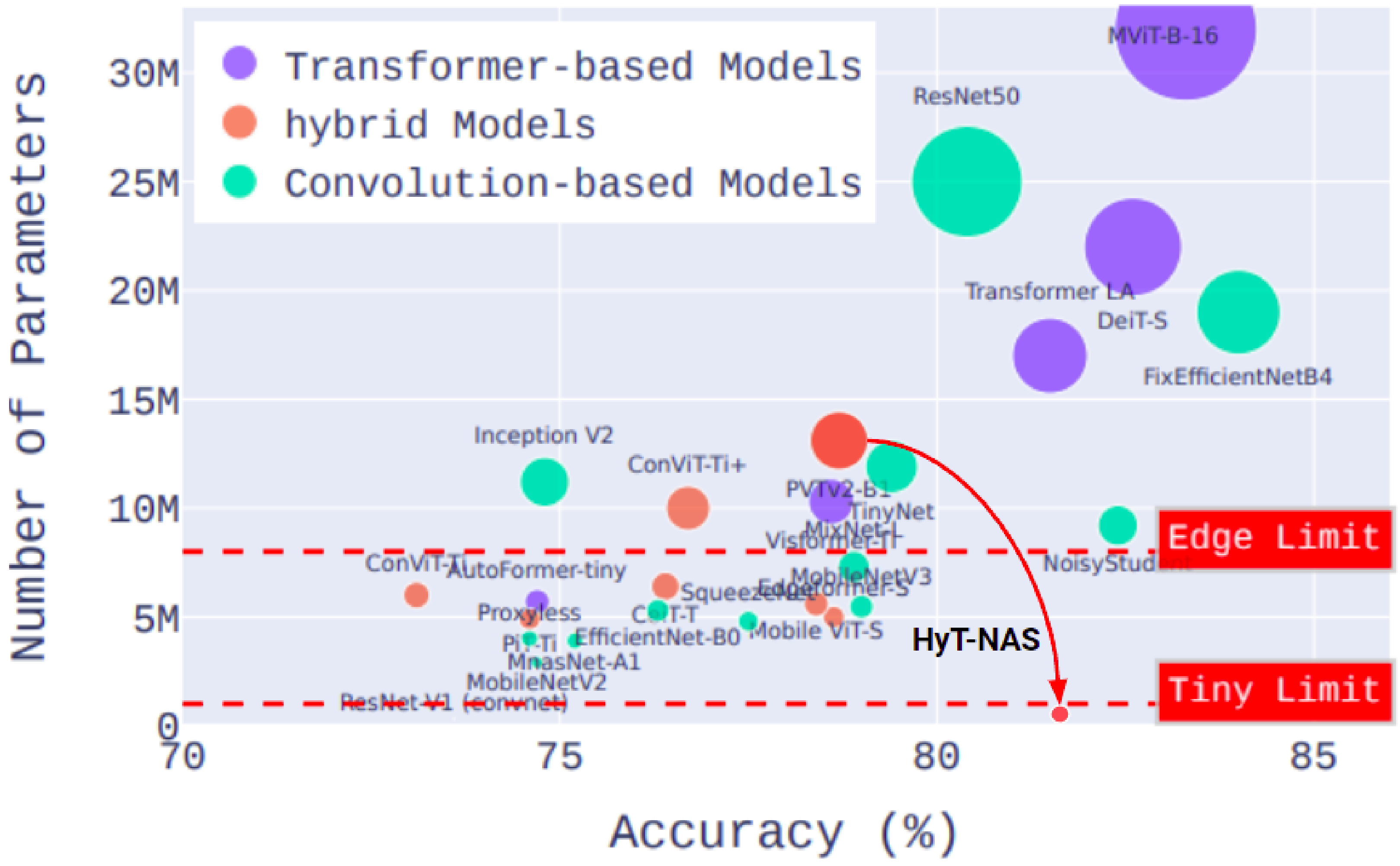
High cost evaluation which consists in training deep learning architectures

Ex: The learning time of ViT on ImageNet1k for 100 epochs on 8 NVIDIA A100- 40GB GPUs is 65 hours.

source: https://ai.facebook.com/blog/significantly-faster-vision-transformer-training/

Objective

Propose an efficient hardware-aware neural architecture search method to find **Hybrid Transformer models** that are fast, deployable on small edge devices and effective for **Visual Object Recognition.**



Study Case

Image

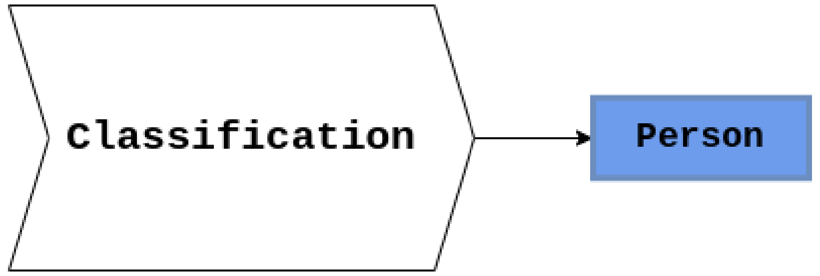


Image Classification

Object Detection

Hardware Aware Neural Architecture Search (HW-NAS)

$$\underbrace{max_{\alpha \in A}\ f_1(\alpha)}_{\text{Accuracy}},\ \underbrace{f_2(\alpha), ..., f_n(\alpha)}_{\text{Hardware efficiency metrics}}$$

A: Defines the space of possible architectures (the hyperparameters considered and its value ranges).

α: an architecture of the space A, defined by the values of its hyperparameters.

Search Space → **Search Strategy** ⇄ **Evaluation strategy**

Search Space

Convolution Neural Networks

YOLO (J.Redmond et al; 2015); MaskR-CNN (k.He et al, 2017)

Vision Transformers

Swin (Liu et al, 2021); PVT (wang et al, 2021); DETR (Zhu et al, 2020)

Hybrid

ConVit (D'ascoli et al, 2021)

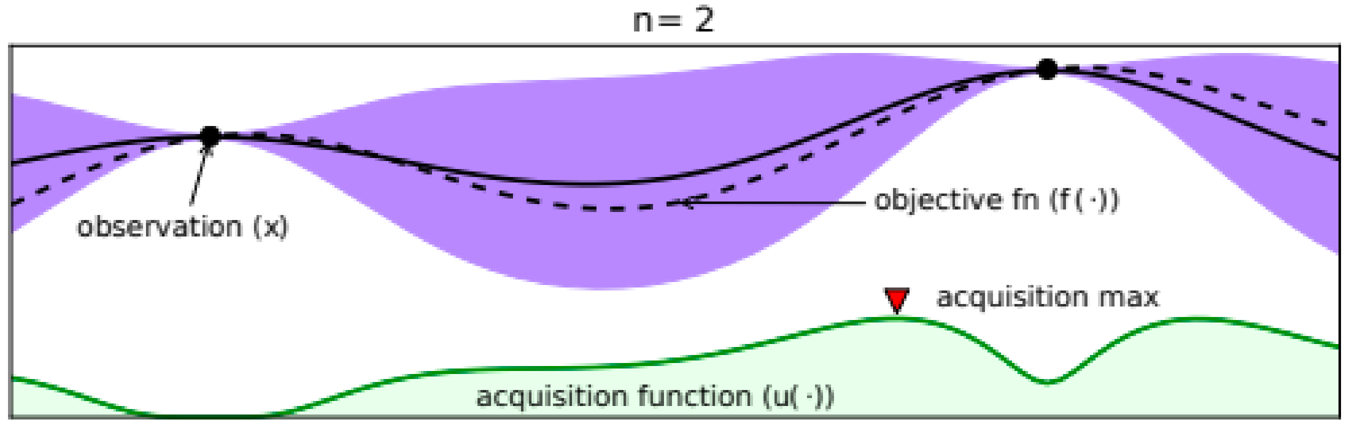
Search Strategy

Compute-based strategies

NSGAII (Deb et al, 2002)

Model-based strategies

Multi-objective Bayesian Optimization (eg: DGEMO by Konakovic et al, 2020)



Evaluation strategy

Reducing the training cost

Small number of epochs (e.g., Zela et al., 2018)
Subset of the data (Klein et al., 2017).

Prediction Models

Predicts an objective such as accuracy (e.g., C. Liu et al., 2018).

One-Shot

Train a single model of the search space (e.g., Pham et al., 2018).

Zero-shot

Zen-NAS (M. Lin et al., 2021): a numerical score used as a proxy for the expressivity of a neural network.

State of the art

Hardware Aware Neural Architecture Search (HW-NAS)

$$max_{\alpha \in A} \; f_1(\alpha), f_2(\alpha), \ldots, f_n(\alpha)$$

Accuracy — Hardware efficiency metrics

A: Defines the space of possible architectures (the hyperparameters considered and its value ranges).

α: an architecture of the space A, defined by the values of its hyperparameters.

Search Space → **Search Strategy** ⇄ **Evaluation strategy**

Search Space

Convolution Neural Networks

YOLO (J.Redmond et al; 2015); MaskR-CNN (k.He et al, 2017)

Vision Transformers

Swin (Liu et al, 2021); PVT (wang et al, 2021); DETR (Zhu et al, 2020)

Hybrid

ConVit (D'ascoli et al, 2021)

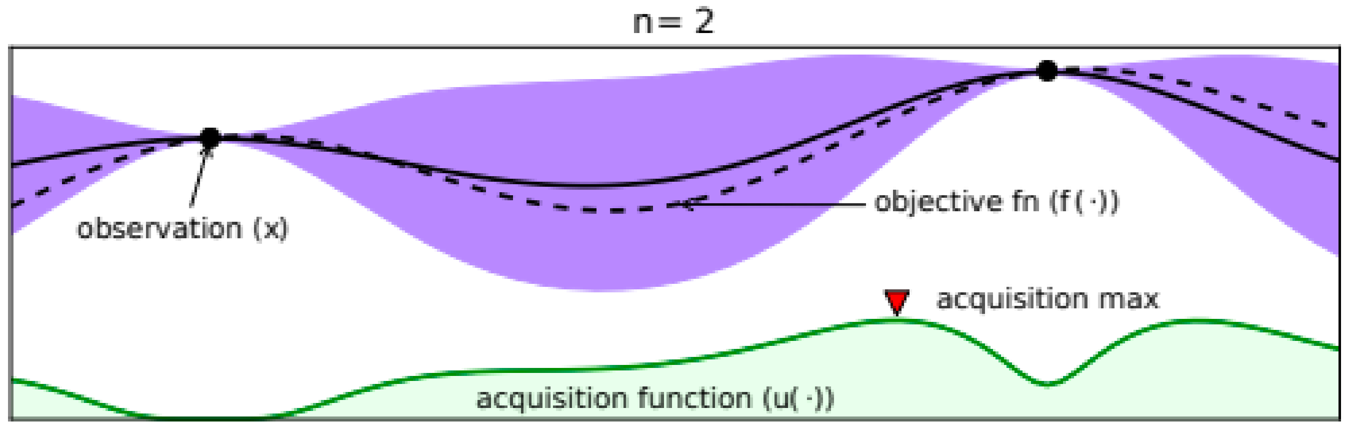
Search Strategy

Compute-based strategies

NSGAII (Deb et al, 2002)

Model-based strategies

Multi-objective Bayesian Optimization (eg: DGEMO by Konakovic et al, 2020)



Evaluation strategy

Reducing the training cost

Small number of epochs (e.g., Zela et al., 2018)
Subset of the data (Klein et al., 2017).

Prediction Models

Predicts an objective such as accuracy (e.g., C. Liu et al., 2018).

One-Shot

Train a single model of the search space (e.g., Pham et al., 2018).

Zero-shot

Zen-NAS (M. Lin et al., 2021): a numerical score used as a proxy for the expressivity of a neural network.

HyT-NAS

HyT-NAS

Hybrid Search Space

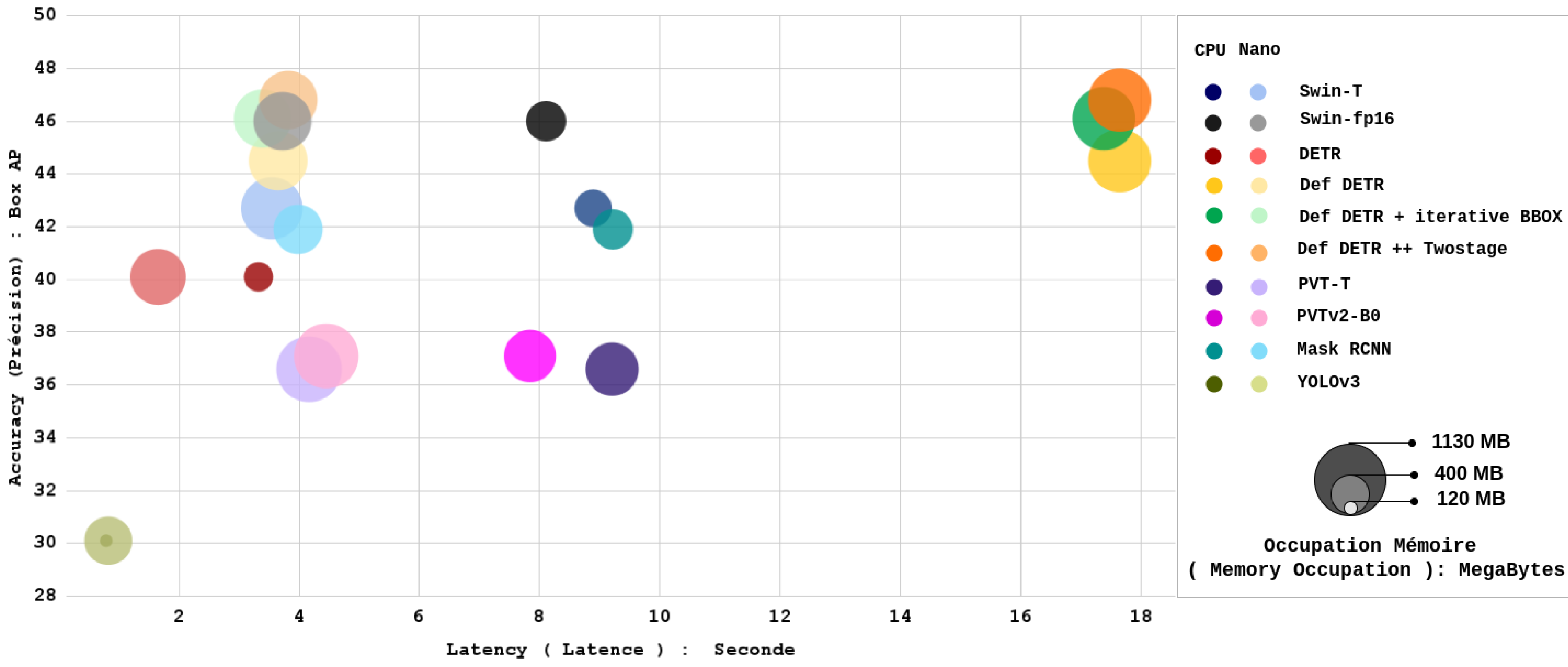
Propose an Initial search space : Accuracy-focused study SOTA architectures for Visual Object Recognition.

- Too big to efficiently explore ~ 10^{27}
- Does not consider hardware constraints

Efficiency analysis : Comparative study of the efficiency of SOTA models and operations on edge devices according to hardware metrics such as Latency, Memory consumption, Size and Throughput.

Hybrid models are more likely to be deployed on edge devices.

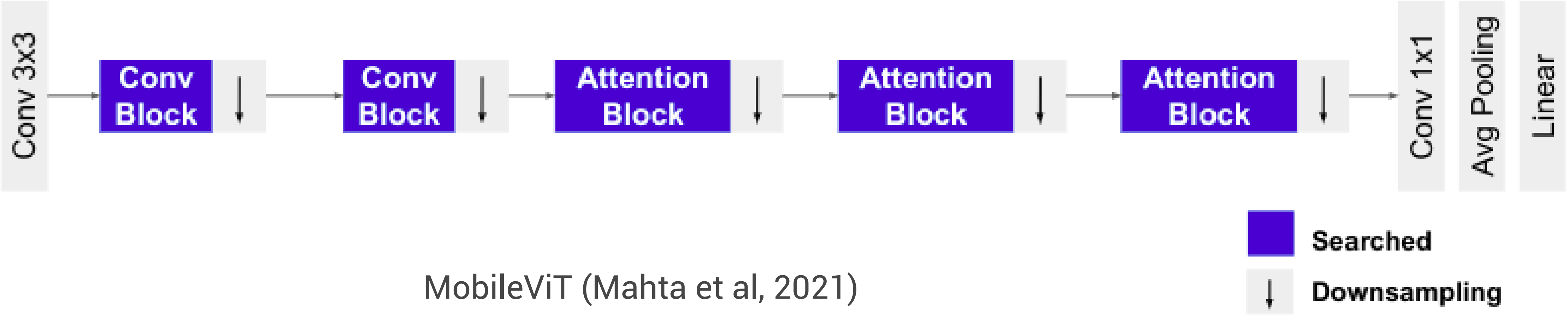
Hyperparameters such as the number of heads and the embedding size have more impact on the size and efficiency of attention blocks than others.



HyT-NAS

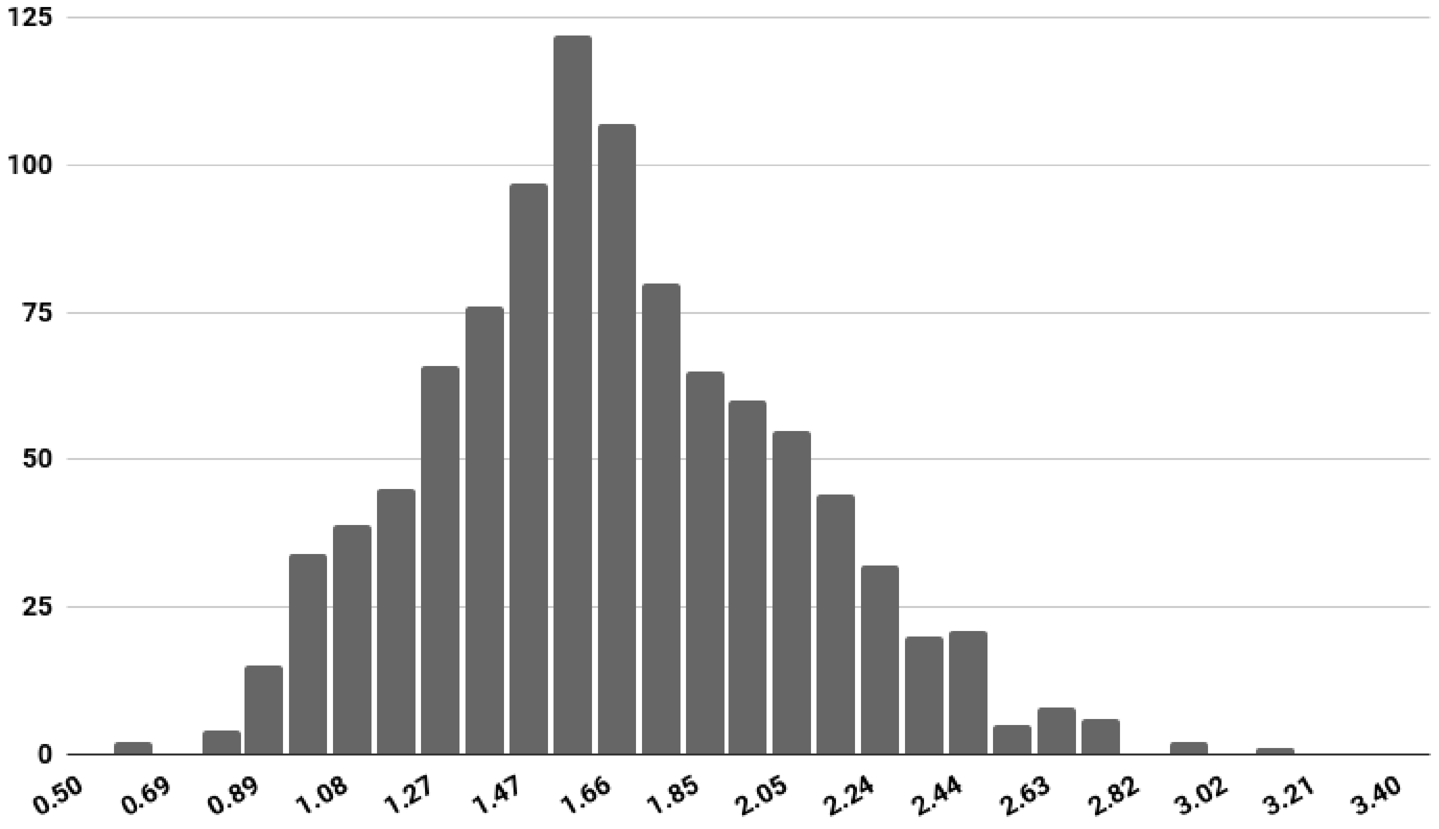
Hybride Search Space

Description



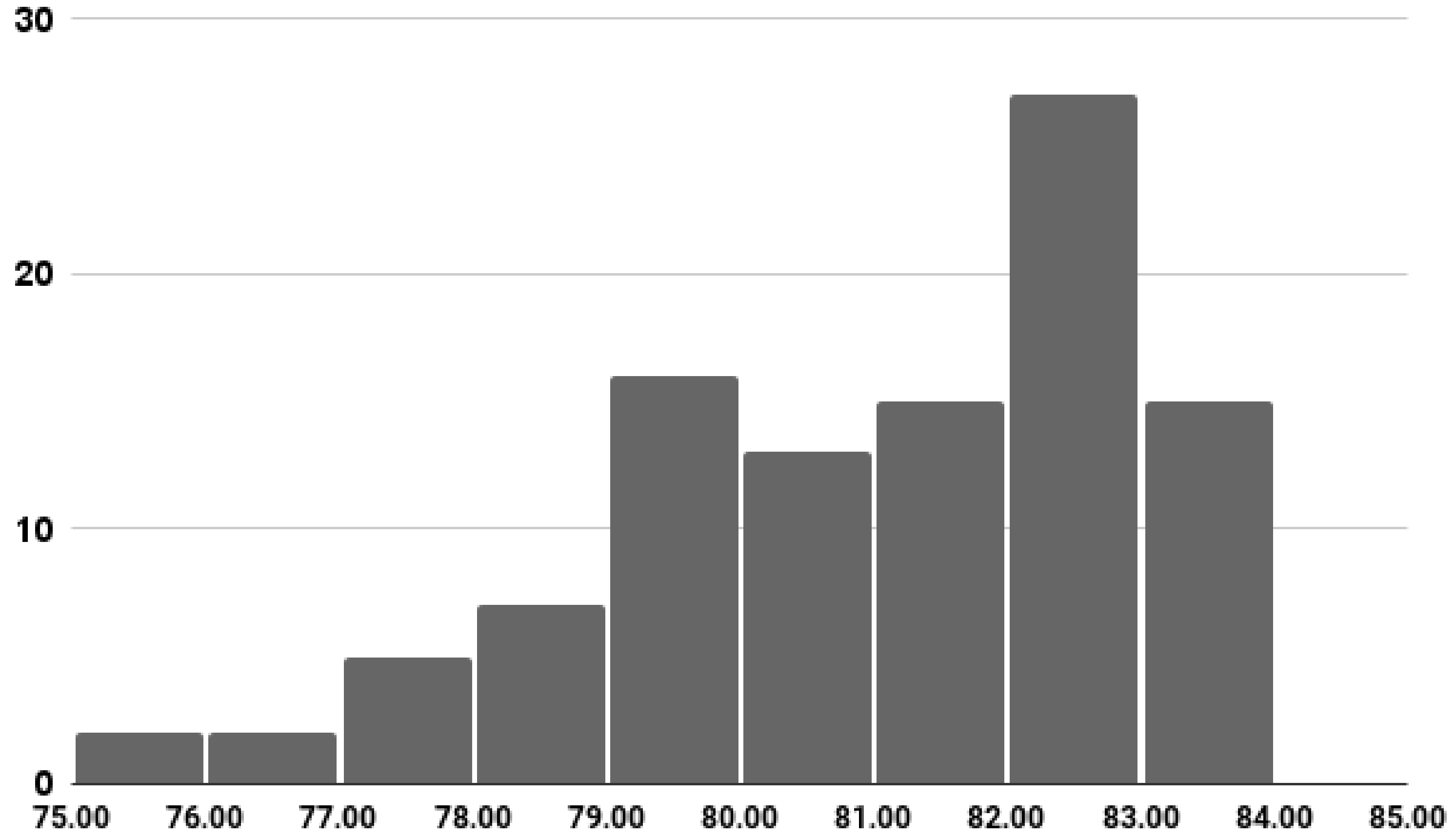
MobileViT (Mahta et al, 2021)

Block	Hyperparameter	Values
Convolution Block	Number of blocks	[1, 2, 3, 4]
	Expand ratio	[1x, 2x, 4x]
	Out channel size	[8, 16, 24, 32]
Attention Block	Expand ratio	[1x, 2x, 4x]
	Channel size	[1x, 1.5x, 2x]
	Number of heads	[1, 2, 4]
	Feed forward ratio	[1x, 1.5x, 2x]



Latency (s)

Evaluation

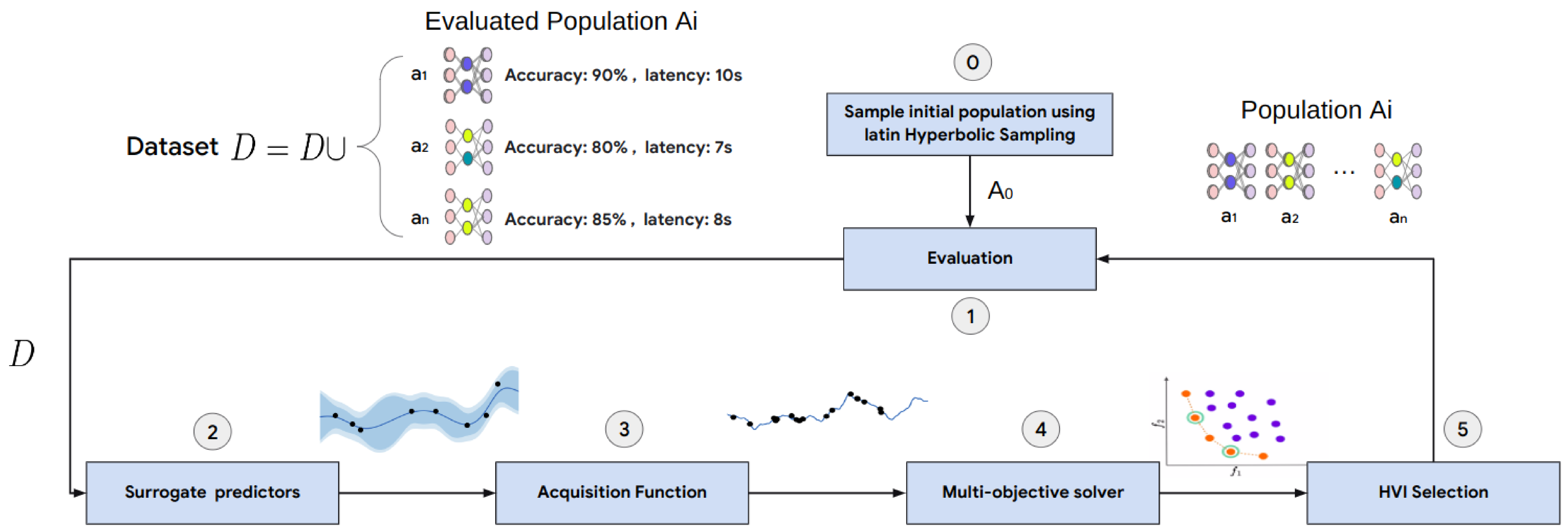


Accuracy (%)

HyT-NAS

Search Strategy

$$max_{\alpha \in HySS}\ Accuracy(\alpha), Throughput(\alpha) \quad subject\ to\ Nparamaters(\alpha) \leq MaxNparamaters$$



fit a set of predictors on all the data D seen previously and obtain the mean and standard deviation.

Assign a score to each probable observation based on its likelihood of helping the optimization process to improve the current solutions.

Solve the optimization problem defined as maximizing the acquisition functions of all objectives.

select, from the output of the multi-objective solver, the best solutions based on the hypervolume improvement.

HyT-NAS

Search Strategy

Study

Surrogate
- XgBoost, XgBRanker
- Feed Forward Networks (FFN)
- Gaussian Process (GP)
- Bayesian Neural Network (BNN)

Acquisition
- UCB (Upper Confidence Bound)
- EI (Expected Improvement)

Multi-objective solver
- NSGAII

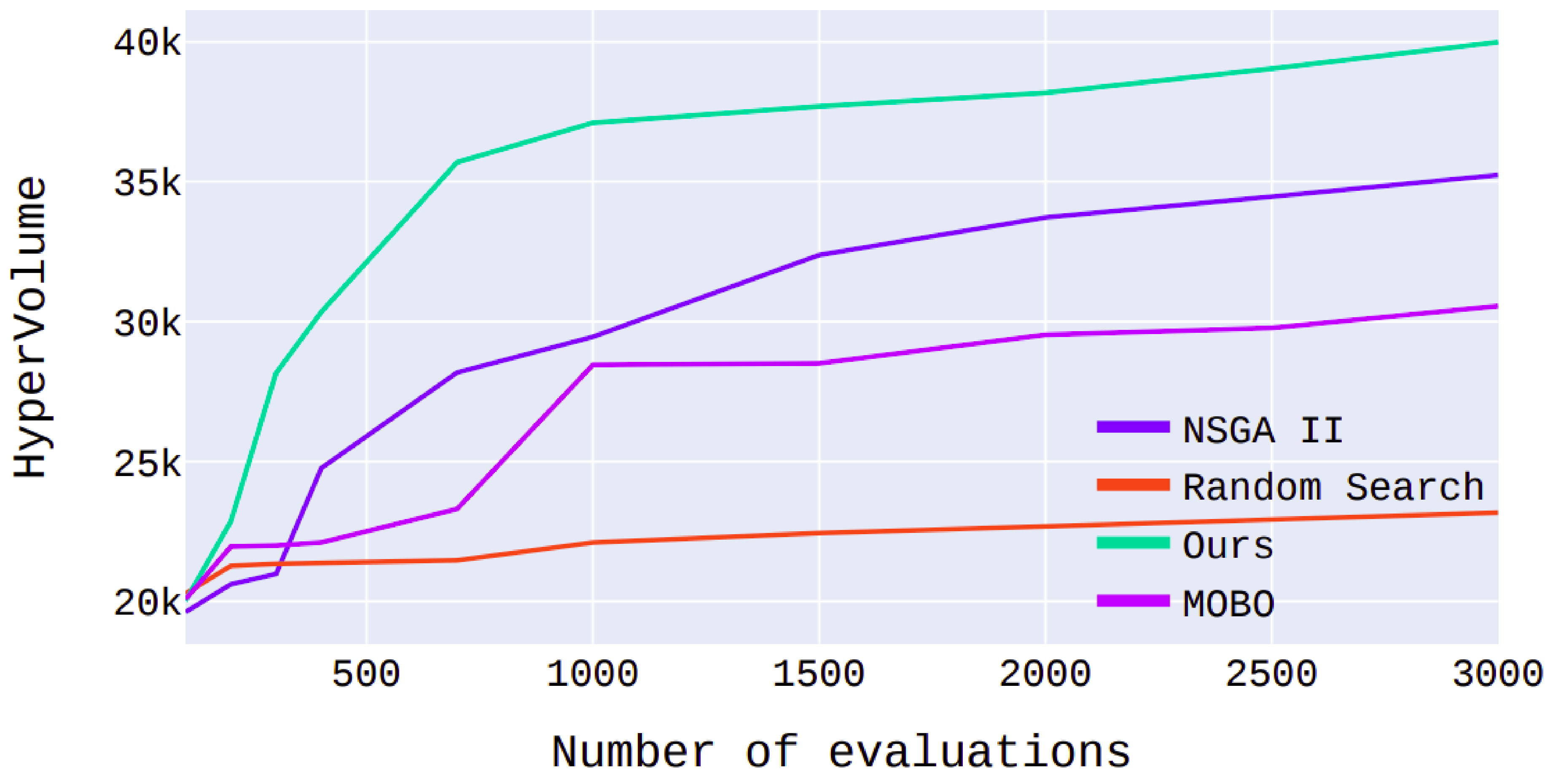
Selection method
- HVI (Hypervolume Improvement)
- Random
- Dominance

Method	Surrogate	Acquisition function	Multi-objective solver	Selection method	Performance (Avg Number of discovered paretos)
Random					3.68/14
CMA-ES					5.45/14
NSGAII					6.06/14
MOBO std	GP	EI	NSGAII	None	5.4/14
HyT-Search	BNN	EI	NSGAII	HVI	5.2/14
HyT-Search	FFN (1layer)	EI	NSGAII	HVI	10.2/14
HyT-Search	FFN(2layer)	UCB	NSGAII	Random	11.4/14
HyT-Search	XGBoost	UCB	NSGAII	Dominance	12.6/14
HyT-Search	XgBoost	UCB	NSGAII	HVI	13.7/14

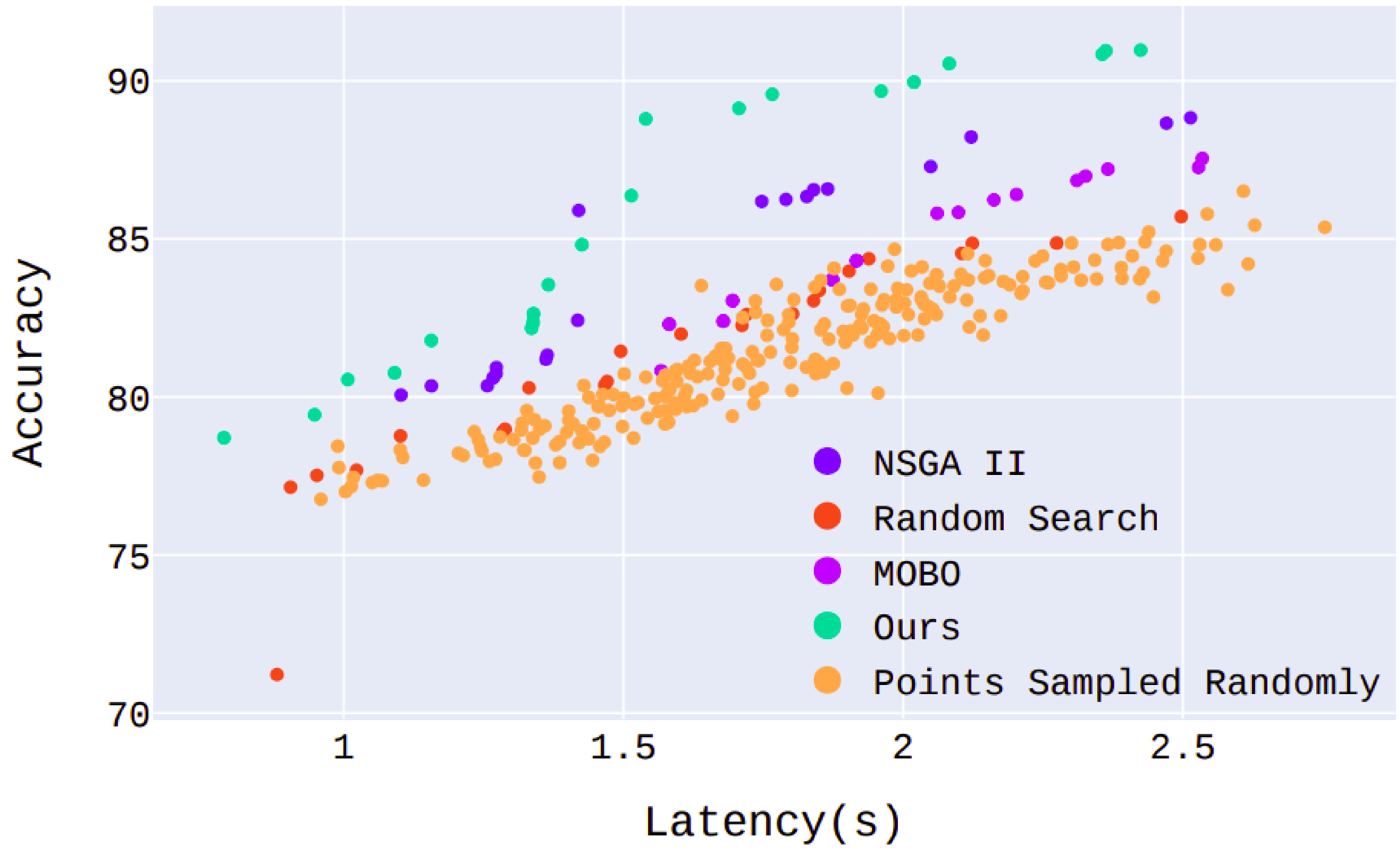
Benchmark: Reproducible and Efficient Benchmarks for Hyperparameter Optimization (https://github.com/Este1le/hpo_nmt)

Search Strategy

Evaluation



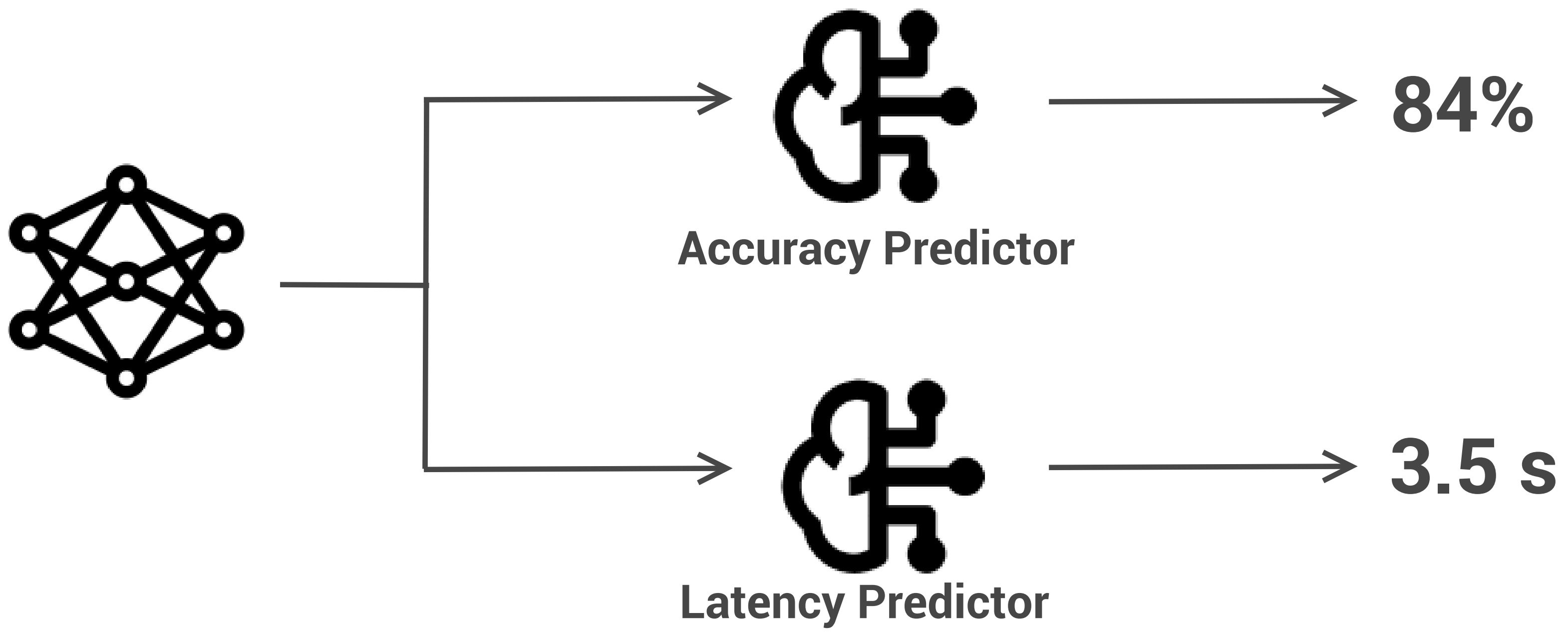
Our method converges faster by obtaining a higher Hypervolume with fewer evaluations.



Our method allows us to obtain better results by discovering a better pareto-front.

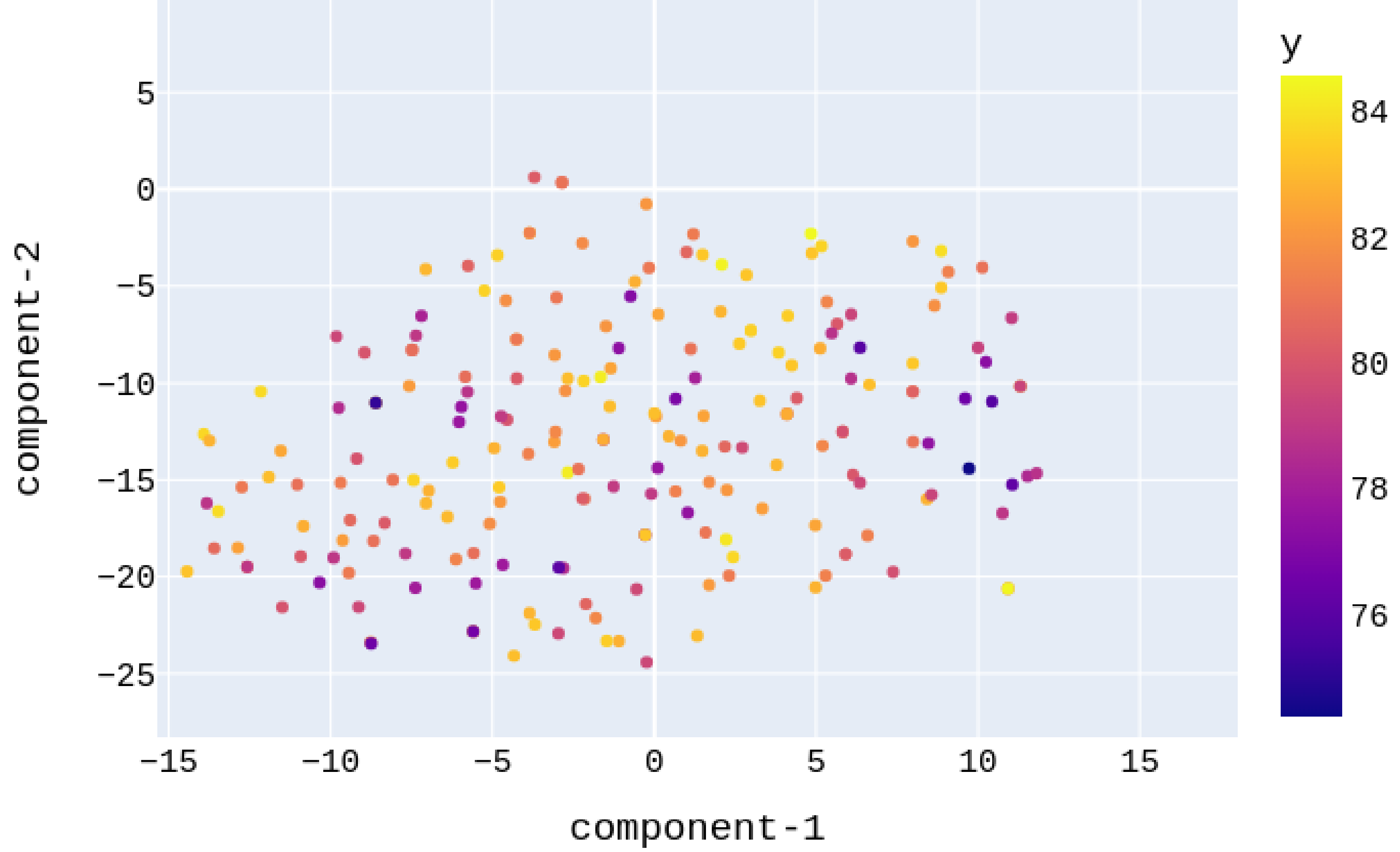
HyT-NAS

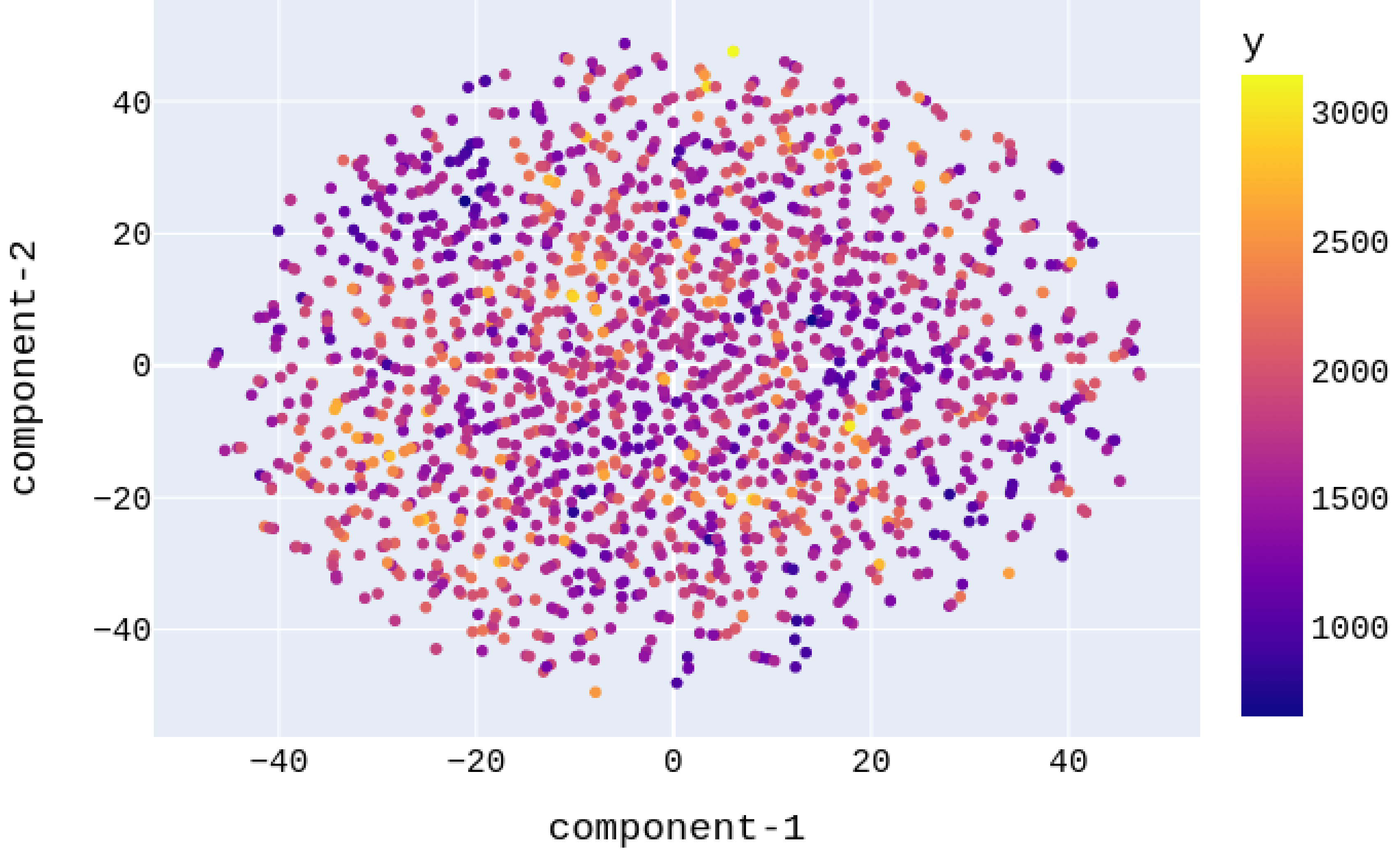
Evaluation strategy



Use predictors of Accuracy and latency to evaluate the selected architectures during the search .

The predictors were trained on datasets constructed by selecting architectures uniformly from the search space and measuring their performance.





Results

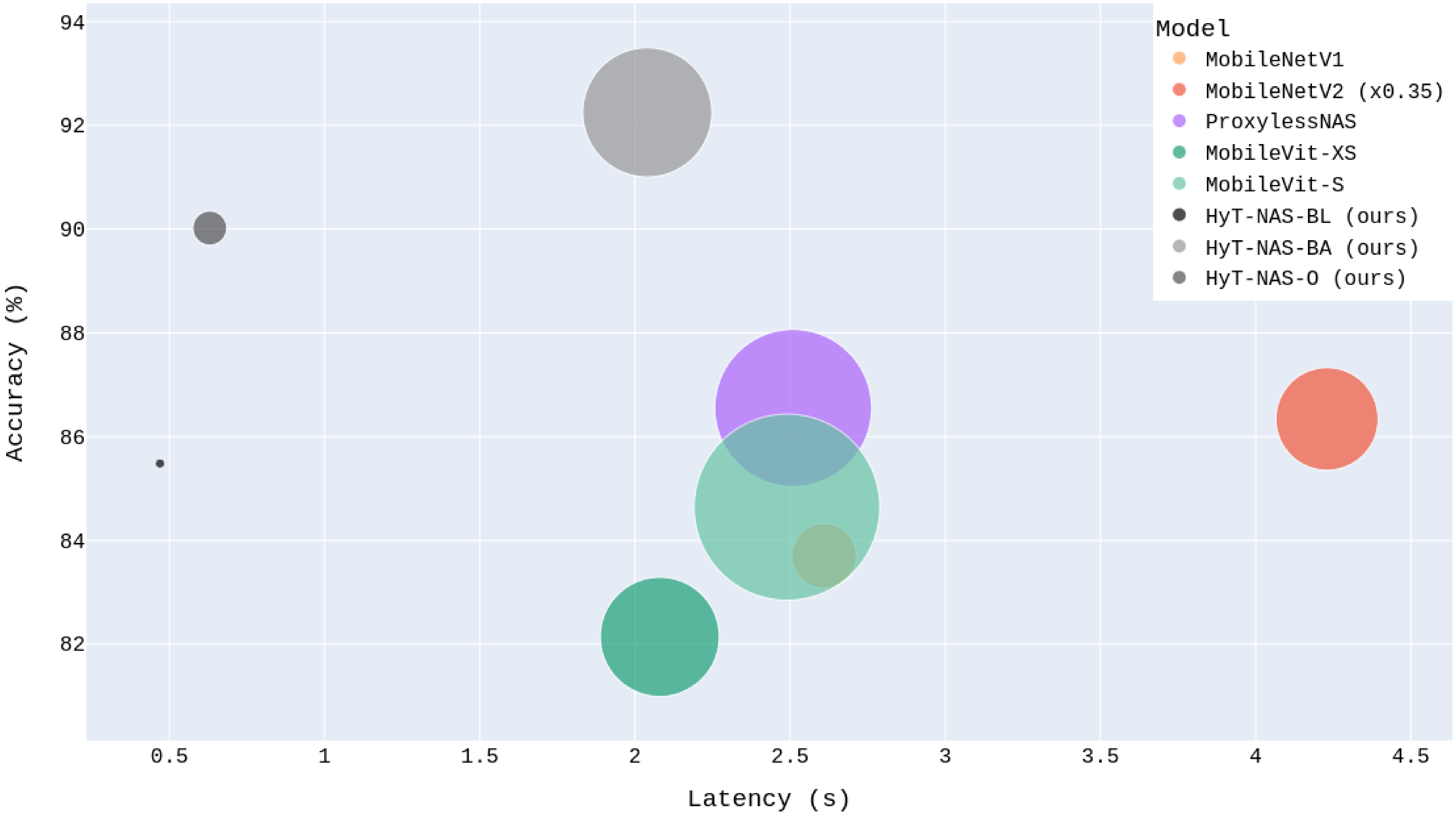
HyT-NAS

Visual Wake Words

HyT-NAS-BL outperforms MobileVit variants while significantly reducing latency and the number of parameters.

HyT-NAS-BA is largely more accurate with lower latency than all the others and a smaller size than the most.

HyT-NAS-O outperforms the 90% in accuracy with a latency and size more optimal than all the others.

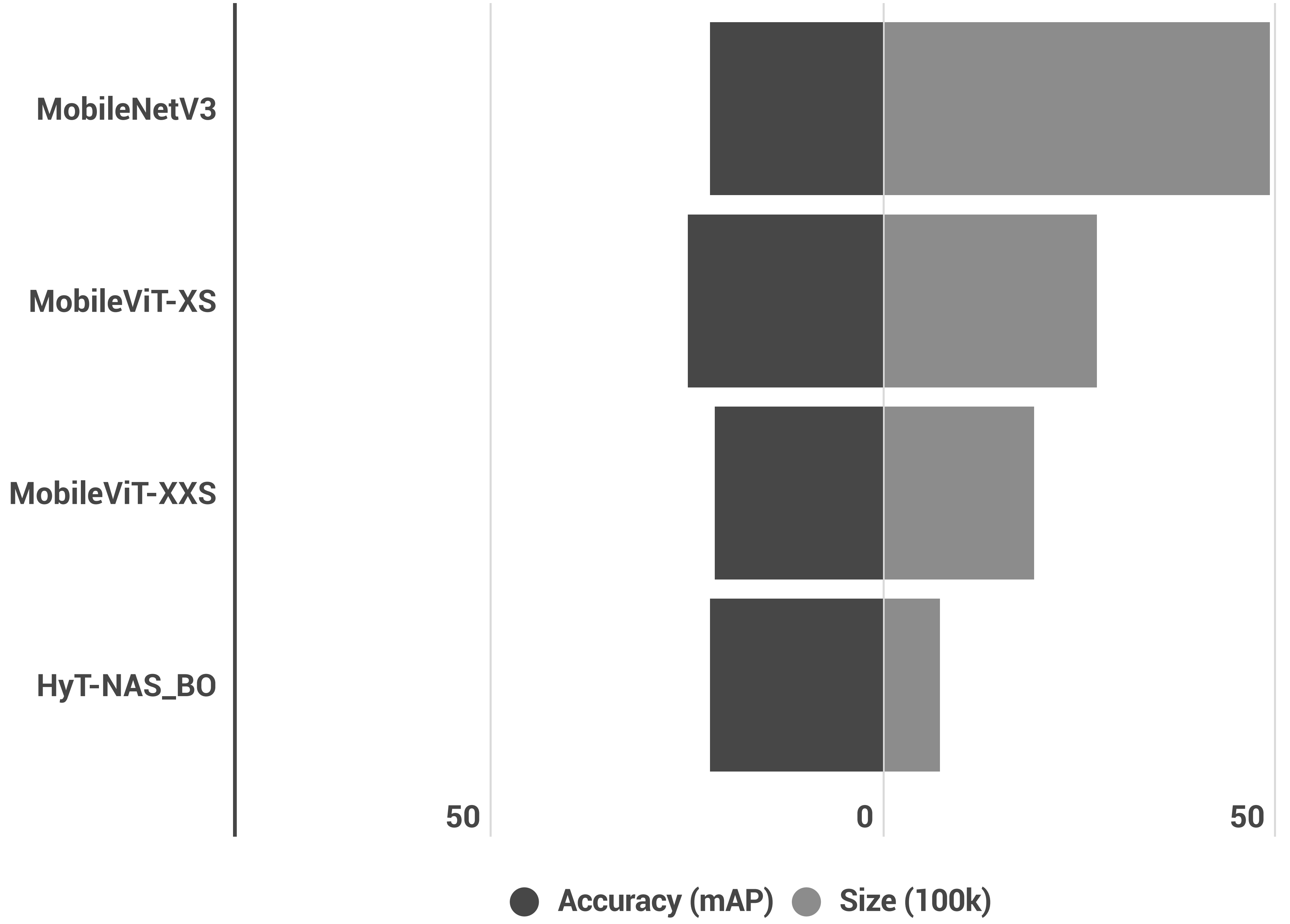


HyT-NAS

Person Detection

> Our HyT-NAS_BO detector achieves better accuracy than mobilenetV3 while being much smaller (more than 5x).

> Our HyT-NAS_BO detector achieves similar accuracy as MobileViT-XXS while being smaller (2.8x).



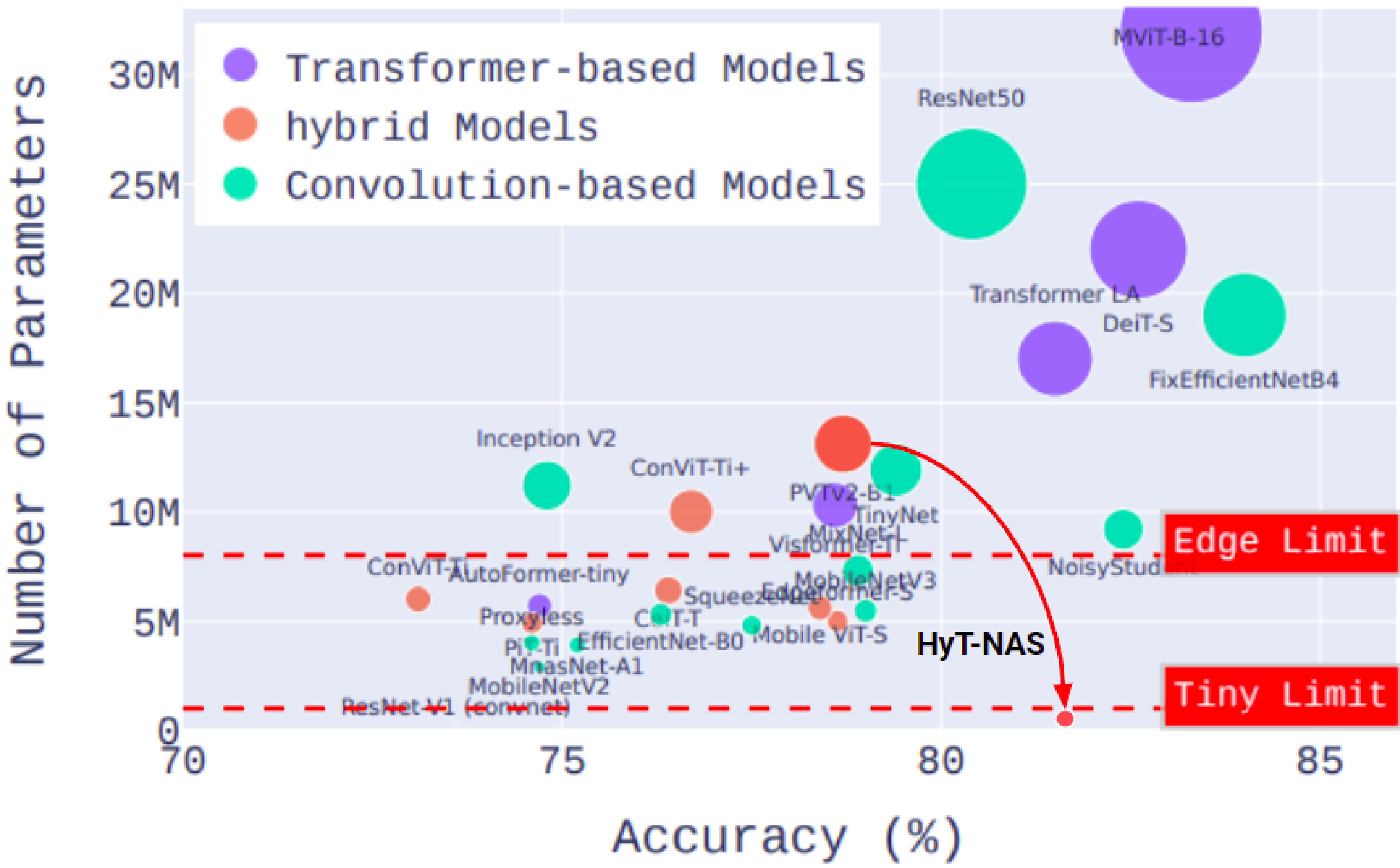
Take-away

Propose a new method of automatic search of neural architecture adapted to the hardware called "HyT-NAS".

Realize a comprehensive study of Vision Transformers models for visual object recognition on several hardware platforms.

Propose a new hybrid search space that includes convolution and attention blocks targeting small edge devices.

Propose a new search strategy aims to accelerate convergence by finding good architectures in a relatively small number of evaluations.



Perspective

Expanded the search space by allowing interchanging of attention and convolution blocks

Consider other metrics in the optimization such as: energy consumption.

Add semantic segmentation as a use case.

Thank you for your attention

HyT-NAS: Hybrid Transformers Neural Architecture Search for Edge Devices

MECHARBAT Lotfi Abdelkrim (ESI ex INI)
hl_mecharbat@esi.dz
https://www.linkedin.com/in/lotfi-abdelkrim-mecharbat-7740b3164/
https://github.com/meclotfi/HyT-NAS-Search-Algorithm