ICPR 2026 · ORAL PRESENTATION

Closed-Loop LLM Discovery of Non-Standard Channel Priors in Vision Models

Can performance feedback turn a LLM into a neural architecture search operator?

Tolgay Atinc Uzun · Dmitry Ignatov · Radu Timofte

Computer Vision Lab · CAIDAS & IFI · University of Würzburg

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

ICPR 2026
International Conference on Pattern Recognition
LYON

Why automate channel configuration?

THE BOTTLENECK

Manually selecting channel widths relies on expert intuition and repeated experiments.

We want to automate this decision and improve it using measured performance feedback.

	MANUAL DESIGN	AUTOMATED SEARCH
PROCESS	Expert chooses widths layer by layer	LLM proposes complete channel configurations
SIGNAL	Intuition and trial-and-error	Measured one-epoch proxy accuracy
SCALABILITY	Repeated redesign for every model	Closed-loop generation and evaluation
OUTCOME	One hand-designed configuration	Evidence-backed channel priors

Why our approach instead of conventional NAS?

THE PROBLEM

NAS search spaces are designed by humans and encode assumptions about which architectures are worth exploring.

An LLM that edits complete source code is not limited to a predefined cell, block, or supernet.

	CONVENTIONAL NAS	CODE-LEVEL NAS
REPRESENTATION	Predefined graph, cell, or supernet	Executable PyTorch source code
SEARCH OPERATOR	Edges, blocks, widths, or continuous relaxations	A code LLM conditioned on code and metrics
VALIDITY	Built into the restricted representation	Must be checked across syntax and tensor dependencies
MAIN TRADE-OFF	Reliable candidates, bounded design space	Broader semantic edits, explicit validity problem

LEMUR and NNGPT

LEMUR · DATA

- ✓ executable PyTorch models
- ✓ Variety of architectures
- ✓ training settings + transformations
- ✓ reproducible performance statistics

Large repository of models and measured outcomes

NNGPT · FRAMEWORK

1 prompt → candidate code

2 validate → train → score

Performance feedback updates the LLM with LoRA and guides the next channel proposal.

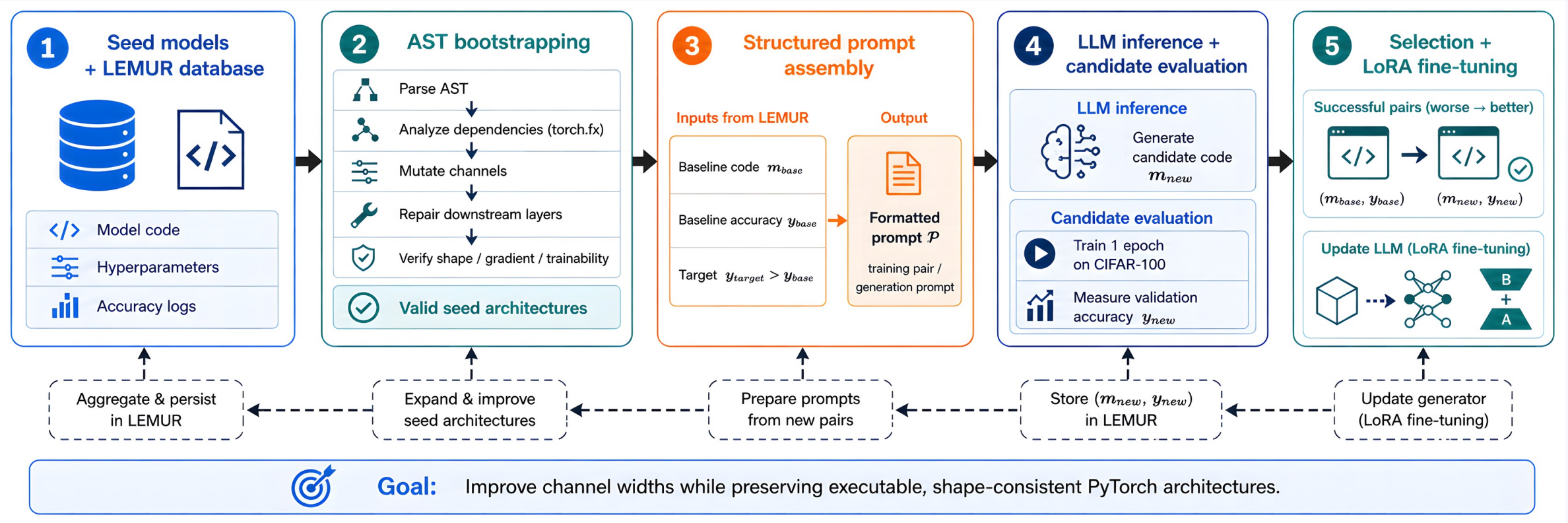
Channel search as performance-conditioned code generation

$$m_{new} = \mathrm{LLM}(\, m_{base},\, y_{base},\, y_{target} \,)$$

INSTANTIATION

TASK	CIFAR-100 image classification
BASELINE CONTEXT	Complete AirNet source code, transform, hyperparameters, and measured accuracy
MEASURED BASELINE	`y_base = 0.250 after one training epoch`
REQUESTED OUTCOME	`Generate an executable candidate with y_target > 0.250 under the same proxy`
RETURNED ARTIFACT	`<hp> optimized hyperparameters  <tr> transformation code  <nn> complete model code`

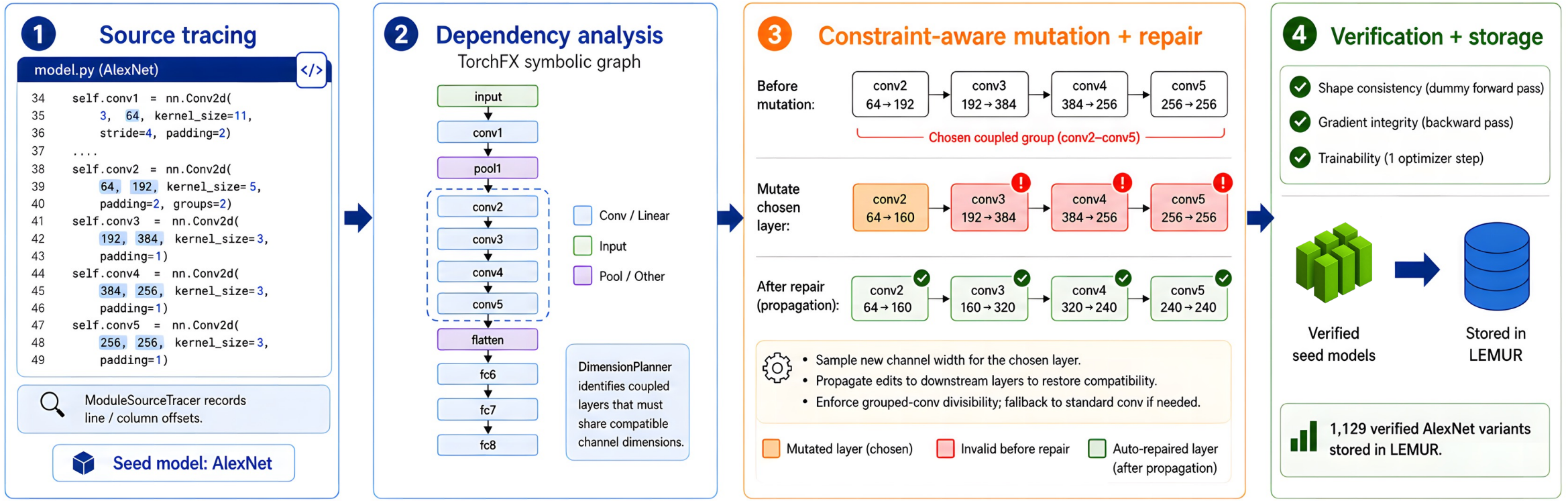
Putting the components together: closed-loop channel-width neural architecture search



Seed data establishes executable structure and the task examples. The measured candidate performance supplies the next fine-tuning signal.

Uzun et al., ICPR 2026, Figure 1

AST bootstrapping produces structurally valid supervision before LLM search



The AST engine is used for initialization only: 1,129 verified AlexNet variants are stored in LEMUR, then the LLM searches an AirNet-based skeleton.

Uzun et al., ICPR 2026, Figure 2

The training prompt schema

SYSTEM

"You are an expert PyTorch neural network architect specialising in channel dimension optimisation."

TASK

"Change the channel configuration of the baseline neural network to improve the value of '{metric}' at least to {addon_accuracy} ... on the '{dataset}' dataset.

GUIDELINES

- adjust channel dimensions strategically
- establish new growth/reduction patterns
- preserve computational and dataset constraints

INPUT

<hp>{prm}</hp>
<tr>{transform_code}</tr>
<nn>{nn_code}</nn>
<metric>{metric_code}</metric>

REQUIRED OUTPUT `<hp>{addon_prm}</hp>` `<tr>{addon_transform_code}</tr>` `<nn>{addon_nn_code}</nn>`

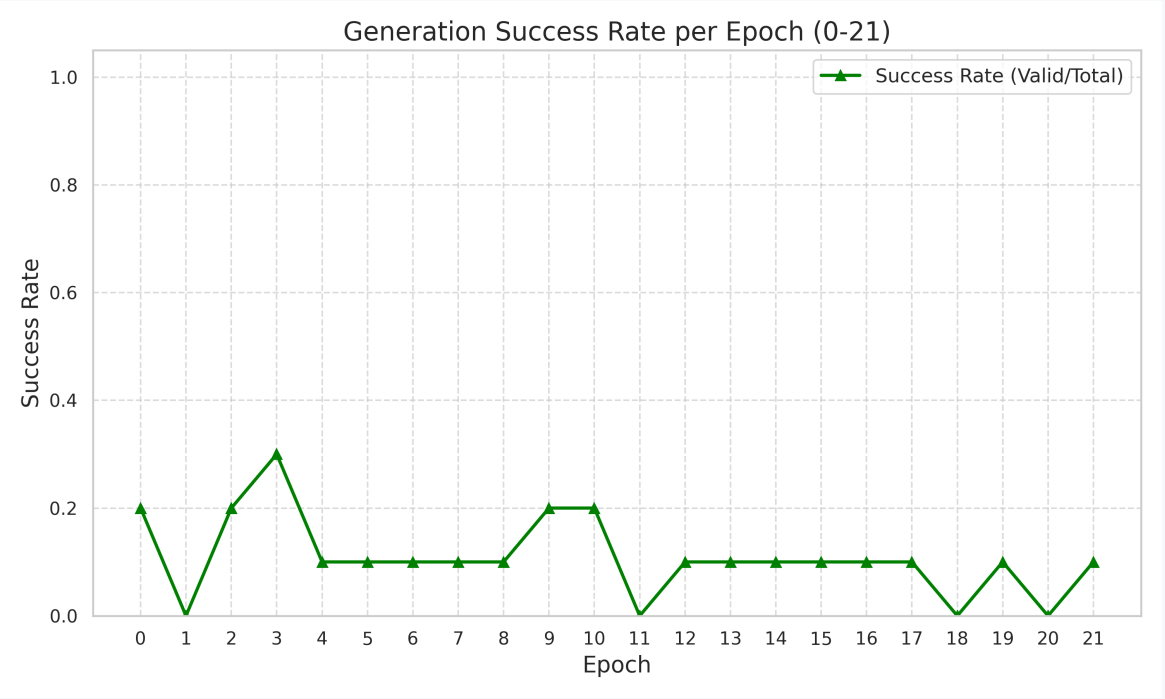
The LLM returns full training configuration and source code, not only a channel vector.

The channel allocation under a fixed one-epoch proxy

SEARCH TASK	CIFAR-100 image classification	AirNet-based skeleton; all convolutional and fully connected widths
CLOSED LOOP	22 cycles (0-21) · 10 candidates per cycle	220 generated proposals in total
CANDIDATE EVALUATION	1 training epoch · batch 64 · AdamW	Fixed augmentation and transforms
GENERATOR	OlympicCoder-7B · context 16,384	temperature 0.8 · top-k 70 · top-p 0.9
LORA UPDATE	r=32 · alpha=32 · dropout=0.05	learning rate 1e-6 · cosine schedule · paged AdamW 8-bit

All reported accuracies are one-epoch proxies, not converged CIFAR-100 benchmark results.

Successfully trained candidates



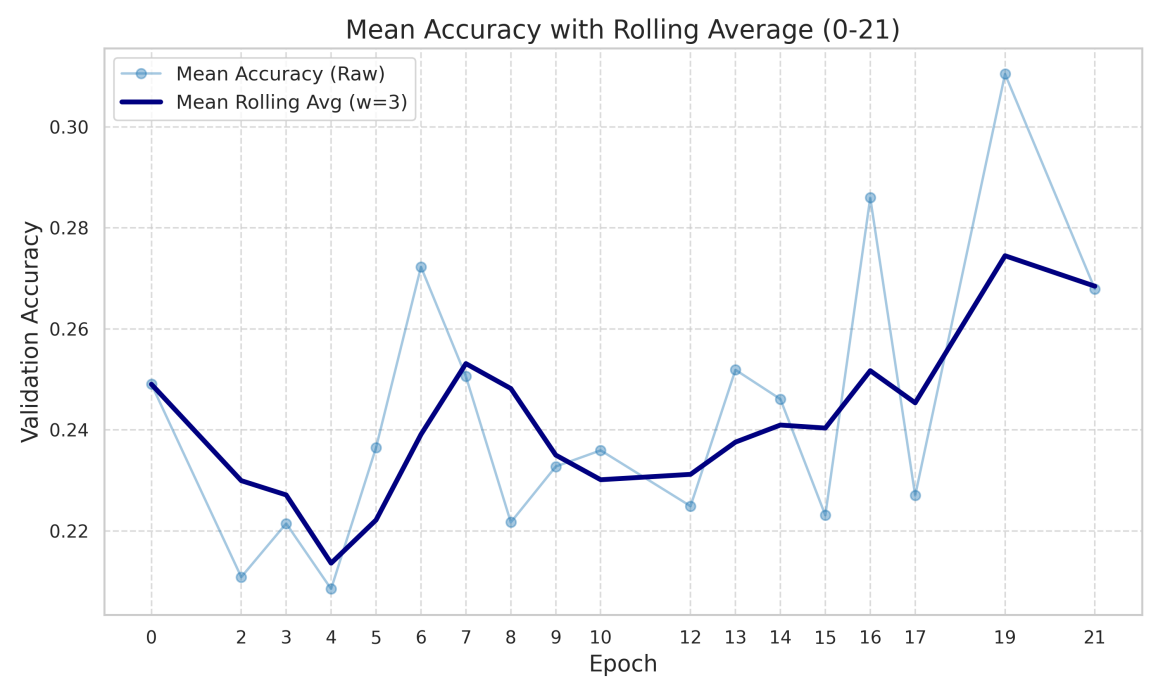
DEFINITION

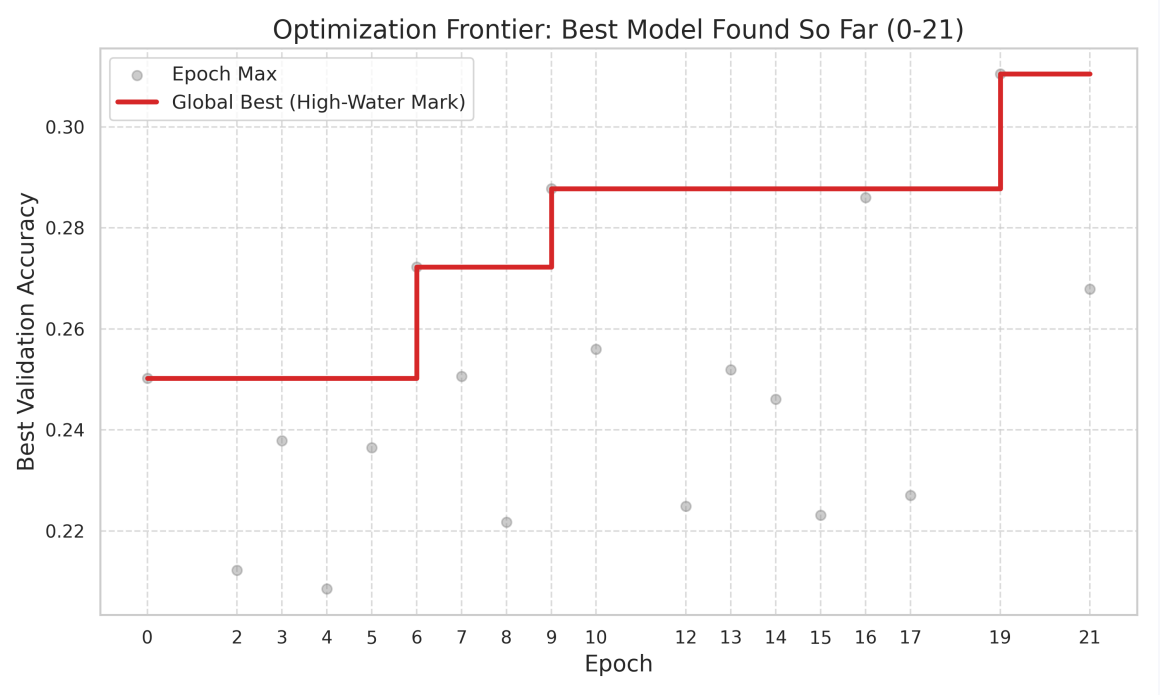
A candidate is valid when it can be trained and evaluated end-to-end; 20 / 220 (9.09%) reach this point.

Invalid candidates are rejected before the accuracy statistics; they affect search cost, not the reported means.

220 -> 200 rejected -> 20 evaluated

Closed-loop search raises the best one-epoch proxy from 0.250 to 0.311





BEST INITIAL **0.250**

GLOBAL MAX **0.311**

+24.1% relative improvement

global maximum at cycle 19

Uzun et al., ICPR 2026, Figure 3 right + Figure 4 right

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

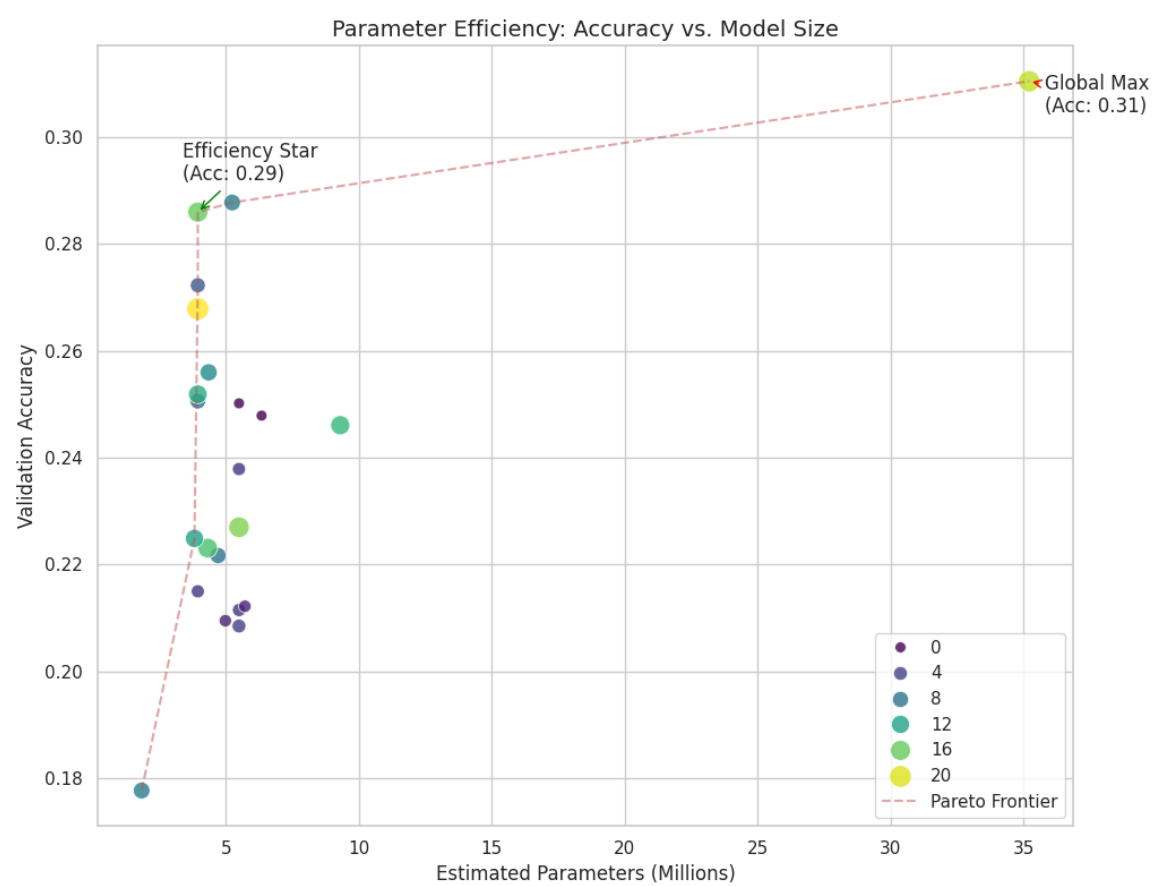
Statistical tests support a positive trend but limited by the sample size

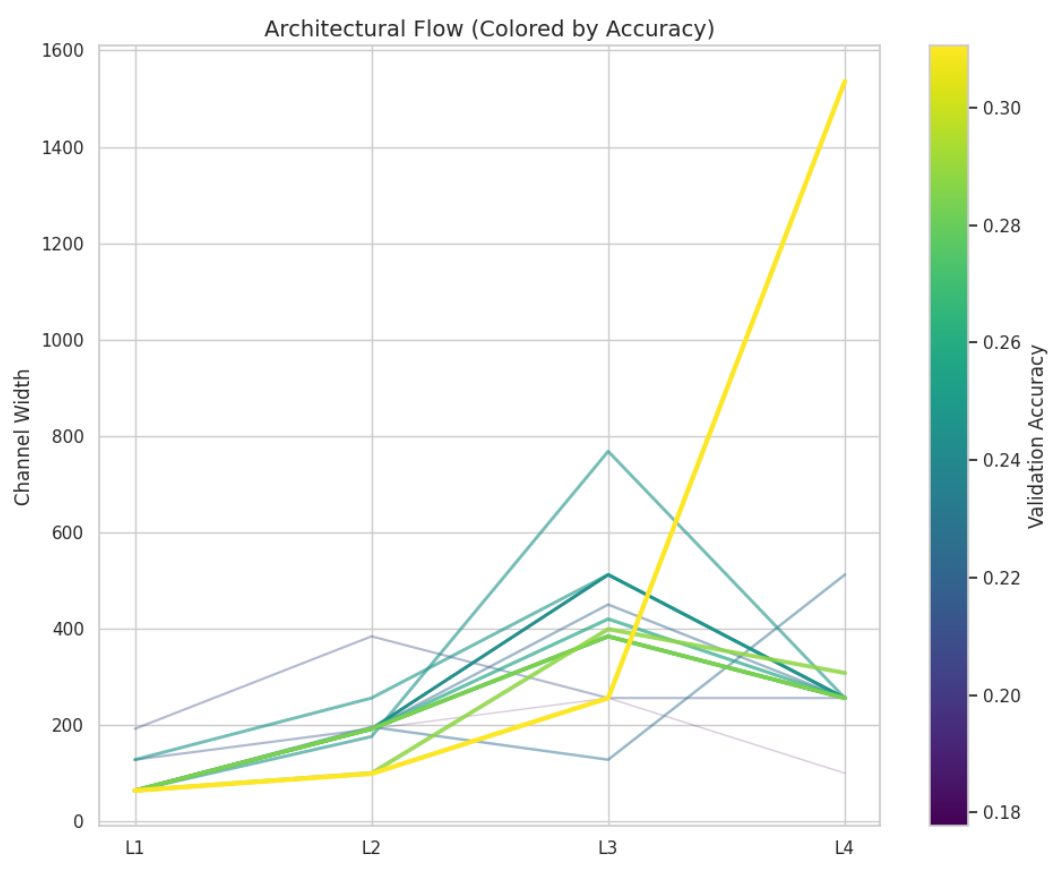
ANALYSIS	ESTIMATE	P-VALUE	READING
Peak-performance regression	+0.0019 / cycle	0.083	Positive trend
Early vs. late valid population	0.226 -> 0.273	0.0033	Higher observed mean in later models
100,000-label permutation test	+0.0474 gap	0.0077	Distribution-free support

SAMPLE SIZES early N=9 (cycles 0-5) · late N=4 (cycles 16-21)

The tests are encouraging, but N=4 limits the breadth of the claim.

The original analysis reveals an accuracy-size Pareto frontier and late-stage expansion





Efficiency was not an explicit prompt objective, yet narrow-early / wide-late candidates appear on the frontier.

CONCLUSION

Closed-loop code generation can support NAS when structure and evaluation are explicit.

DATA	LEMUR supplies executable models and measured training context.
METHOD	AST initialization plus NNGPT feedback produces performance-directed supervision.
RESULT	The best proxy improves by 24.1% and exposes channel priors.
SCOPE	One dataset, one backbone, one-epoch proxy; no random search comparison.

github.com/ABrain-One/NN-GPT

Tolgay Atinc Uzun · Dmitry Ignatov · Radu Timofte

Questions

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG