BICEC: Attachable Classification-Based Intelligent Control for Sustainable Computer Vision Systems

Jonathan Burton-Barr [1 2], **Deepu Rajan** [1], **Basura Fernando** [2 1]
[1] NTU College of Computing and Data Science, [2] A*STAR Centre for Frontier AI Research

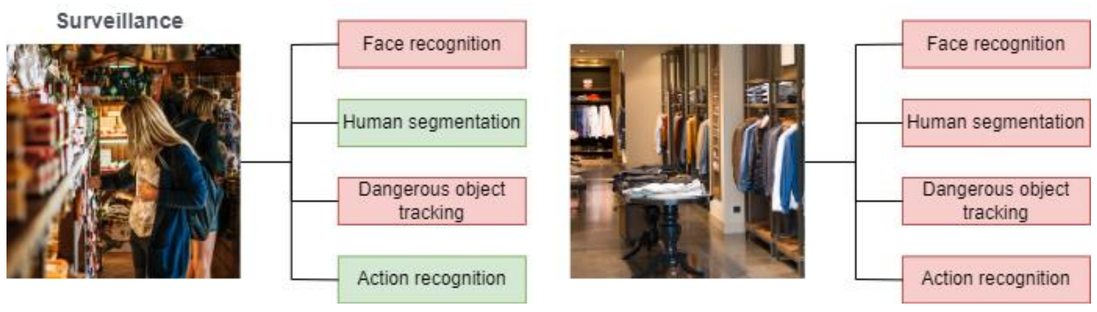
Motivation

Sustainable AI: Identifying irrelevant inputs enables input skipping, leading to reduced computational load and energy consumption
Rising inference costs: The accumulated and individual inference costs of AI is increasing
Modular systems: Fine-grained control over system components can improve efficiency, flexibility, and sustainability

System Control via Intelligent Methods

- Intelligent control has proven effective in reducing resource usage in non-intelligent systems
- It is important to regulate **multi-model AI systems**, where redundant processing creates excessive consumption

- **BICEC** (Binary Image Classification Evaluative Controller) and its predecessor **SICEC** [1], initiates a line of research into **intelligent redundancy reduction**
- **BICEC**: An attachable controller which learns **activation conditions** → condition present → model receives input



Task and Data

Input Relevance Task (IRT): Human-centric task with four base functions (IRT-B) and two extended functions (IRT-E)

We reuse datasets from the system pre-trained models. Since BICEC uses binary, independent classification, each branch only requires **positive and negative examples**.

Training

Phase 1 Base creation
Split EfficientNet-B0 blocks into 1–5 (shared base) and block 6 (branch-specific). Load pre-trained EfficientNetV2-B0 weights. During training:
- Branches update their own layers independently
- Shared base updates via combined loss

Phase 2 Branch adaption
Set a tolerance threshold (α) for accuracy drop. For each branch:
- Find minimal width/depth maintaining accuracy ≥ (baseline – α)
- Initialize scaled branch with weights from Phase 1 (uniform element selection [3])

Branch addition
Integrate new branches for new system vision models:
- **Step**: Train new branch (shared base frozen)
- **Pull**: Freeze new branch, update shared base using combined loss of all branches

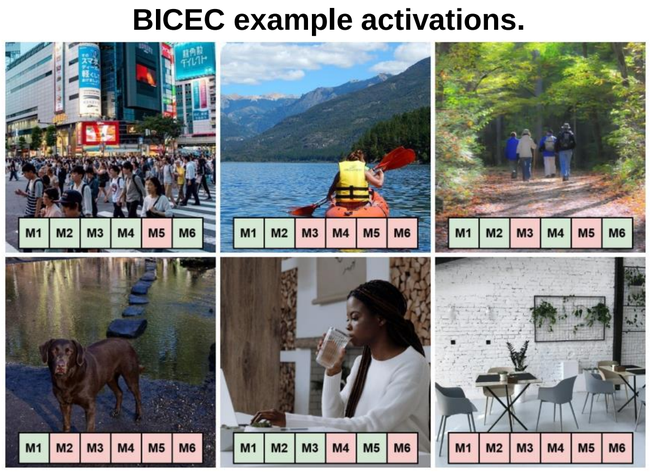
Future Changes

- Switching to vision transformer backbone
- Expanding the rule set
- Generalized model relevance
- Testing more multi-model systems

Citations

[1] Burton-Barr, J., Fernando, B., & Rajan, D. (2024). Activation Control of Vision Models for Sustainable AI Systems. *IEEE Transactions on Artificial Intelligence*.
[2] Tan, M., & Le, Q. (2021, July). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.
[3] Xu, Z., Chen, Y., Vishniakov, K., Yin, Y., Shen, Z., Darrell, T., ... & Liu, Z. (2023). Initializing models with larger ones. *arXiv preprint arXiv:2311.18823*.

BICEC example activations.



IRT Functions and Activation Conditions

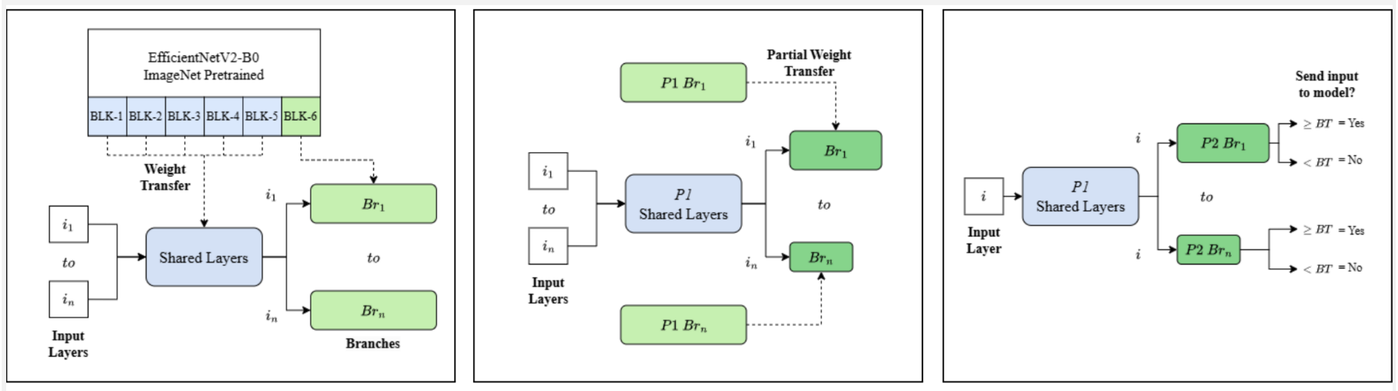
Ref	Function	Activation Condition
M1	Object Detection	Animates
M2	Segmentation	People
M3	Face Detection	Faces
M4	Pose Detection	3+ People
M5	Action Recognition	Call, Text, Eat, Drink
M6	Segmentation	Clothing Accessories

AI Vision Systems

1. We may wish to use multiple domains
2. Available models may collaboratively be able to perform the AI system requirements
3. It may be difficult to create a single model that encapsulates all functionalities
4. Function-specific modulation may enhance system performance

Branched Binary Classification Network

- **Backbone**: Built on EfficientNetV2 [2] – lightweight design prevents bottlenecks; optimized for both width and depth scaling.
- **Activation conditions**: BICEC learns when model-specific conditions are present in the input.
- **Branched architecture**: Enables precise learning of individual model conditions and forms the basis for BICEC's adaptive, expandable design.
- **Binary threshold**: Tunable after training to control model activation sensitivity.

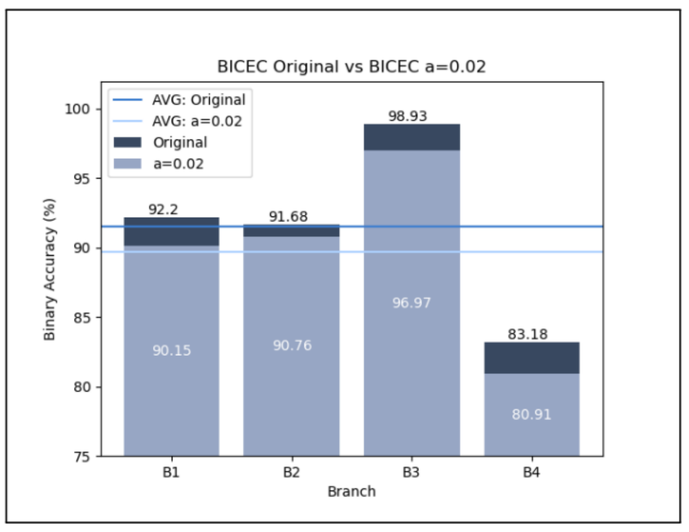


(a) *Phase 1*: Base Creation. (b) *Phase 2*: Branch Adaptation. (c) *Final Model.*

Main Results

Correct and Incorrect Model Activation (CMA / IMA): How often a BICEC branch correctly or incorrectly activates a model

Branch accuracy before and after branch adaption.



BICEC model size and GFLOPS after scaling: Increasing α enables lower BICEC size which can alleviate system bottlenecks.

a	Br_1	Br_2	Br_3	Br_4	Model Size	GFLOPs	Acc (%)
Orig.	1.0	1.0	1.0	1.0	19.50M	2.591	91.50
0.005	0.9*	0.9	0.9*	0.3	13.03M	2.055	90.90
0.010	0.9	0.7	0.7	0.2	8.39M	1.706	90.54
0.015	0.9	0.6	0.3	0.3	6.55M	1.520	90.02
0.020	0.4	0.4	0.3	0.2	2.47M	1.181	89.70

Adjusting binary threshold: Setting a lower binary threshold improves the chance that both correct and incorrect models will be activated.

BT	CMA	IMA	Accuracy
0.5	86.43%	8.16%	89.70%
0.25	90.98%	15.15%	88.45%
0.125	93.26%	21.06%	86.60%
0.0625	94.55%	27.42%	84.01%

Branch addition results: Accuracy (A), Network (Net), Branch (Br), Branch after scaling (Br-S). We observed branch addition also improved the performance of BICEC's prior branches. We note that accuracy, CMA, and scale (and thus overall network size), as with Br_1 to Br_4, are strongly influenced by the activation condition.

	Scale	ΔParams	ΔGFLOPs	Net A	Br A	Br-S A	Net CMA	Br CMA
IRT-E T5	0.2	+99.2K	+0.01	+0.12%	90.76%	89.55%	-0.22%	90.00%
IRT-E T6	0.1	+10.2K	+0.003	+0.66%	99.39%	99.09%	+0.79%	99.39%

Average energy and inference costs per model: This is the average cost per input, standard shows the IRT vision systems without BICEC. For our IRT we found a total model energy consumption reduction of **52.1%** and total inference time reduction of **54.7%**.

	Inference (ms)						Energy (W)					
	M1	M2	M3	M4	M5(E)	M6(E)	M1	M2	M3	M4	M5(E)	M6(E)
Standard	17	33	13	12	46	20	1.18	0.86	0.45	0.88	0.38	0.67
COCO-Val	**9.1**	**17.1**	**3.6**	4.1	**8.5**	6.0	**0.63**	**0.45**	**0.13**	0.30	**0.07**	**0.20**
Movie	10.5	20.0	9.1	**2.3**	24.6	7.4	0.73	0.52	0.32	**0.17**	0.20	0.25
Y-VLOG	10.9	22.6	7.5	4.3	17.9	**5.9**	0.76	0.59	0.26	0.31	0.15	**0.20**

Individual branch attention plots: We see each branch focusing on different but activation condition relevant information within the image.