



MixAtlas: Uncertainty-aware Data Mixture Optimization for Multimodal LLM Midtraining

Bingbing Wen^{1,2}, Sirajul Salekin¹, Feiyang Kang³, Bill Howe², Lucy Lu Wang², Javier Movellan¹, Manjot Bilkhu¹

¹Apple, ²University of Washington Seattle, ³Virginia Tech



Motivation

Subhead

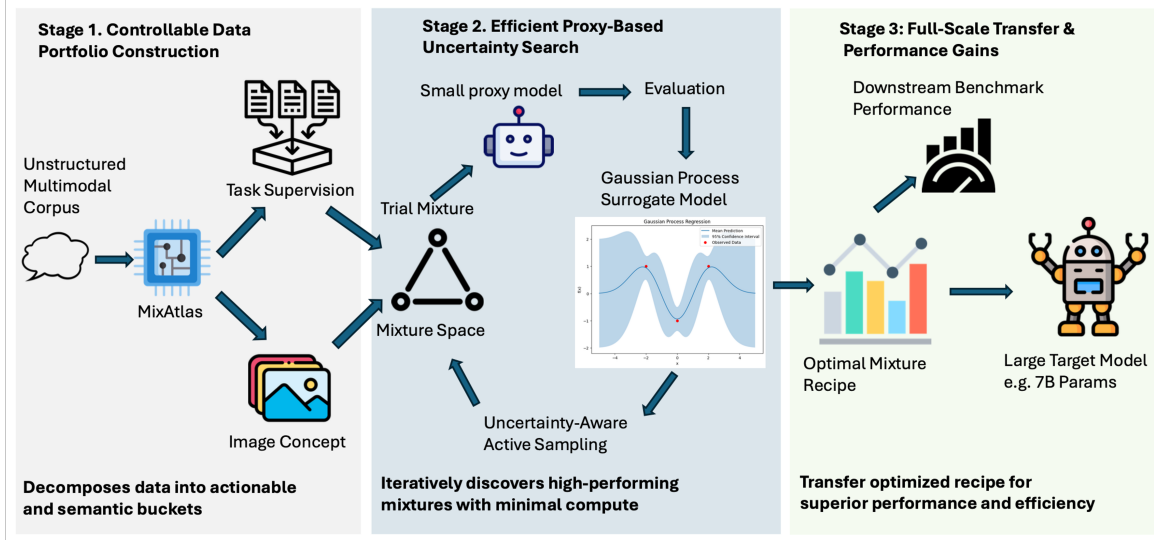
Data mixture optimization for multimodal LLM midtraining remains underexplored. Current approaches rely on simple heuristics or tune mixtures from only a single perspective (data format or task type).

Objective

Subhead

Can we build an interpretable, compute-efficient framework for multimodal mixture optimization that factorizes data along multiple perspectives, discovers recipes via proxy models, and transfers them to full-scale training?

Overview



Results

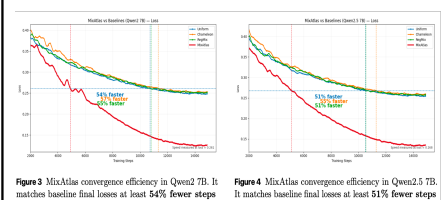
MixAtlas recipes outperform strong baselines at 7B scale

Benchmark	Task Axis				Image Concept Axis			
	Unif	Reg	MixAtlas	Gain	Unif	Reg	Cham	MixAtlas
Visual Understanding								
Ai2D	58.16	58.10	59.84	+2.0%	54.30	53.00	56.50	57.20
CQA	18.95	20.90	39.65	+46.6%	25.14	23.35	23.90	36.48
ScienceQA	72.53	72.30	74.81	+3.1%	71.60	72.90	72.30	71.50
Document and Text								
DocVQA	60.17	58.30	62.98	+4.7%	3.50	15.80	3.00	46.90
TextVQA	50.03	48.10	54.51	+9.0%	29.90	30.00	25.60	50.90
ChartQA	43.44	46.00	50.16	+9.0%	33.30	38.20	34.60	41.80
Multimodal Reasoning								
MMBench-EN	63.40	63.50	67.35	+6.1%	53.78	68.39	59.54	61.89
MMMU Val	39.78	39.80	41.67	+4.7%	37.40	40.00	38.10	40.30
MM-Vet	28.12	26.10	27.75	-1.3%	27.39	26.15	23.21	30.37
Mathematical								
MathVista	32.60	31.80	37.10	+13.8%	35.59	33.40	36.20	34.50
AVERAGE	46.72	46.50	50.68	+8.5%	37.19	40.12	37.30	47.18

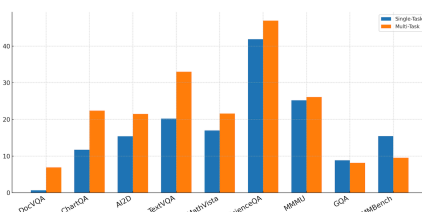
Recipes transfer across model families and scales

Benchmark	Task Axis				Image Concept Axis			
	Unif	Reg	MixAtlas	Gain	Unif	Reg	Cham	MixAtlas
Visual Understanding								
Ai2D	58.65	60.80	59.97	-1.4%	53.31	62.21	58.16	57.19
CQA	32.61	39.00	45.05	+15.5%	25.14	44.27	23.90	36.48
ScienceQA	74.62	73.80	73.18	-1.9%	74.07	73.53	73.33	71.49
Document and Text								
DocVQA	56.03	66.40	74.31	+11.9%	22.68	24.05	25.86	46.83
TextVQA	53.24	56.70	64.24	+13.3%	33.10	48.76	40.60	50.92
ChartQA	52.84	53.50	53.96	+0.9%	38.40	41.16	39.08	41.88
Multimodal Reasoning								
MMBench-EN	60.40	72.80	61.68	-15.3%	63.23	65.89	62.97	60.82
MMMU Val	41.11	42.00	42.67	+1.6%	38.67	42.56	39.00	40.33
MM-Vet	31.38	22.60	31.51	+0.4%	25.60	27.16	27.89	31.42
Mathematical								
MathVista	37.50	37.80	36.10	-4.5%	33.90	38.00	34.90	34.70
AVERAGE	49.84	52.54	54.27	+3.3%	41.01	46.76	42.57	47.21

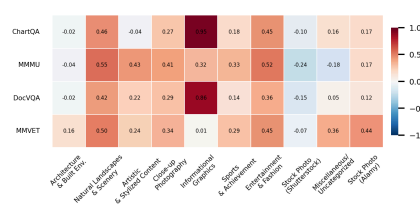
MixAtlas substantially improves training sample efficiency.



Multi-task mixtures surpass single-task midtraining



Benchmark performance is highly domain sensitive



Benchmark performance is highly task sensitive

