

- **Question:** Supervised/Zero-shot names known only to experts, such instead ask foundational VLMs to recognize these
- **Preliminary Results:** they may *not* directly address this challenge



- observations



Democratizing Fine-grained Visual Recognition

Ground-trutti. D		Ground-trutti.		
BLIP-2 P	^{ekBble} fiscovereen Names	BLIP-2	TerrBLIP-2 Terrier Quantiva	tive Results
LLaVA D	ogLLaVA Dog	LLaVA	DogLLaVA Dog	
LENS C	chihuahua Chihuahua	LENS	Brown Dog	Pokemon-10
MiniGPT Shinv ^{SI}	TWOMA ⁴ Chicken ^o Hawk Gerfalcor	Pika Garris	Puppy	$\alpha \wedge CC = \alpha \wedge CC$
FineB'(Ours)	rutimeR((Otirs), Brussels Griffon	FineR (Ours):	Airecine (Ours): Airedale Terrier	CALC SALL
OT, TYLO, TH	edrake, rokeweed, Archean Eor	1, Panouriiorii Ground-truth:		70.0 00.0
			Colordula	/0.8 89.2
BLIP-2:	Source Pikachu			
				34.6 33.1
eqing Pika	Chu, Turtle, Plant, Pokemon, Bu	Ibasaur, Bird		32 3 55 4
				52.5 55.7
Ground-truth	Bentley Continental GT Coupe 2012	FineR (Ours):		39.2 55.7
Ground trutha D	aperestric him hapanese chim, Squirtle,	ware cound truth:	Wheaten Terrier	
BLUP:2210000	tto #Pichu Sadtail Pikachu. Flow	er Seurietle	TerrBLIP-2 Terrier	
BLÍP-2:	Bentley Continental GT	LLaVA	Shabba Vaog F113 Gagy (Ggurs)	/0.8 81.0
CaSED:	Bentley Dog	LENS	DogLENS Dog	
Eine D (Oure)	Bentley Continental CT Seden 2010	MinigP1-4	Shaiyuyi CBJ-4	
Finer (Ours):	Benney Commental GT Segan 2010	FineR (Ours):		
Ground-truth	: Black-and-tan Coonhound	Creand truth		
WordNet:	Black-and-tan Coonhound	Ground-truth:	Squarteuno-trut	NER
		BLIP-2		C Brouged
BLIP-2:	Doberman Pinscher		TUTLELAVA	
CaSED:	Coonhound Irrow	LENS		
FineR (Ours):	: Black-and-tan Coonhound	MiniGPT-4	TurtMiniGPT-4	
Cround truth		FineR (Ours):	WarkineR Ours):	
Ground-truth				