



## Uncovering the Inner Workings of STEGO for Safe Unsupervised Semantic Segmentation

## Alexander Koenig, Maximilian Schambach, Johannes Otterbach

Merantix Momentum, Berlin, Germany

{firstname.lastname}@merantix.com

	Cocostuff				Cityscapes				Potsdam			
ethod	Unsupervised Cluster probe		Supervised Linear probe		Unsupervised Cluster probe		Supervised Linear probe		Unsupervised Cluster probe		Supervised Linear probe	
	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU
EGO (theirs)	56.9	28.2	76.1	41.0	73.2	21.0	_	_	77.0	_	_	_
EGO (ours)	√ 56.9	<b>√</b> 28.2	<b>√</b> 76.1	<b>√</b> 41.1	<b>√</b> 73.2	<b>√</b> 21.0	89.6	28.0	<b>√</b> 77.0	62.6	85.9	74.8
NO (theirs)	30.5	9.6	66.8	29.4	-	_	-	-	-	_	-	-
NO (ours)	<sup>†</sup> 42.4	<sup>†</sup> 13.0	†75.8	<sup>‡</sup> 44.4	52.6	15.2	<sup>‡</sup> 91.3	<sup>‡</sup> 34.9	71.3	54.3	84.5	72.8
CVPR #***** nission #*** <b>CVPR 20F20 EN1</b>	5hAis <b>5i5i∕iæ</b> ₩*C.C		le 1: Vali	dation re	CVPR Suits*of	reproduc	ibility st	udy.				
160	CVPR		CVPR	CVPR	CVPR	CVPR CVP	R CVPR					
iessteliditienittoexperimenti Southe fieles applientienti 2011	talshalvdationFtg	guensureThe seem applyeby applying	næmtati Figueæd2 gurð xelpaontebyu	pr <b>Ttesseg ment</b> at byoapplyingeal×	tion head proce	sses each #	163 Manalii	na STE	GO's V	Vorkino	ı Princ	ciples
e performs	each other i	nat is, neighbor n this head. Ale	engrapaturestos oheotleer ihethis with films	s, noeigmboenge i ohelædpetVlorreove lefekingetrekteliger	entures do not er, the module	performs sæden øcesses ea	165 166 162	.9				
	DELETER DELETER				ndesing as beat tout	solo an free	163 168 164	hoad Q .	radurae 4	dimoneia	nality f	ഹന
	NTY SONE SECON				tospercipites places pl	intot Actor into	1768 fr	Dotroo	= 90  for	Cocosti	iff. The	k-Mear
		SDANC than ore SUMPLIES IN SUM				sphaces The authorized and the second s	$\frac{167}{167}$	verges b	etter in Ic	wer dim	ensions	s due to
				Cases inclusion of the second	THIS SHITS BEFE	ised of CU	<sup>73</sup> se of di	mension	ality.			
		BUKANDAN MY	sentyinhte Dieso	tosquitoche ed by	patterns that ar the DINO back	e already 171 bone and 172	175 171 176 172		J			
				Principality in the second sec	iende armeen segister proves	tabiensegmentati ellfpssiger Poet	<b>nesis:</b> 78	STEGO is	s a semai	ntics-pre	serving	
	i ulcesta parese int	Buildese about sur		iten heet nig bilest Abjeite Killingspilest	divide differencie al anti-	<b>Tenation</b> National and the second se	onality	reductio	on techni	que spec	cifically	suited
				lizietiska lingetiet		paint an head op	ns style	clusterir	ng algorit	hms		
Alex Alex	feyaganghan Bin	esamplingsampli			<b>bertigreperide</b>	<b>uq</b> i <b>me</b> rmediate i	182 178 P 183 179		ainad aur	ava in <b>Eia</b>		adiaata
	tecture 1st given					noblet Can	Gumulai 185tontial	for dime	aneu cur	ve in <b>rig</b> tv reduct	ion tecl	nuicales
Field Reconcertaine	e Netrie Same net state state Before being					ne egne artig sipieraised traini				ly reduct		Inques
				ne water garver	PT HEIP LEBYSTY LISU HEIPSE THE PROPERTY LISU	stamplanty of t	188 184 VO 189 185					
				instaarteiddige Instaarteiddige	nesideskeps mersielt izvide chilaele olga	Here pairs	180 190 186					
				<b>rigin theatign the</b>	<b>nsitizens de militari</b> y <b>Astrizens de militari</b> y	geto a co- lig 1.0	191 192					
				nikali je esitesettis Usite ile cirbinte pie	in the main of the second s	sbondi <b>a</b> g	193 194					
				arity of by they	calculate the co	From head $0.8$	195 196					
				into Science Speciate		afin a fea-	197 198					
					The subscription of the second	Hentation 0.6	199 200					
		CHURCHUS CONCERNE CHURCH STOLES Includes Handles Saught the Upsam		statuteprobabili no badi	depenin sonielatach i		201					
	SOCTOR TOMOLOGIC		atheomesitentee	Durgenge de durest	chored the second of the secon	<i>⊈wij</i> <u>4</u> 0.4	202				Citvso	apes
204 205 205 205 205 205 205 205 205 205 205	S CHASE SHE AND THE	TELEVISION STREET	Des ann rencry rier	<b>itaritization des Ekte</b> i S <b>ploen Serice Co</b> nfeat	ture correspond	$\frac{1}{10000000000000000000000000000000000$	204				Potsd	am
See Postal American 206		sholes, which	can the interpret	A swaitchroappe	interpreted as	a form of	206 207					stuff
		insesianhesaso	gattivpositissente,k	ient paic another p	ositiveteken4pa	air and to- Arategies 0.0	208 209					
					HEDESTOR HASSP	Attial cen-	210 211 <b>()</b> D <sub>ViT</sub>	$D_{ m ViT}$	$D_{\lambda}$	/iT	3D <sub>ViT</sub>	$D_{1/iT}$
							212 <u>16</u>	4			4	- vii
do the green ared name ad	coamous no fose p	essitives and nega	Nice Charposel		sematic pressia	wards the From the	213 214	Nur	nber of (	Compone	ents	
n <b>nd is set Thicken Clove 15</b> [10 D <sub>STEGO</sub>	$S_c$ <b>activation</b> $N_c$	nction outsal, the lbs	ss fusæstikbmed nrif F	terson Copester	<b>sor C<sub>STEG</sub> 1.2</b>	Pairs: The	215 <u>~1</u> ~					
	igiczsteiestik	obiotiéEtteskéhneiksEithy sebijystiszteiteenduditi	itaihteiseter võitebat 19 asteriads Microrti	<b>helje)coolite,vlag</b> dsp <b>he</b> p <del>essitiin Ereich</del> n	t <b>hin</b> vi(acdest)bcoat th <b>in</b> vi(acdest)bcoat	ng tignal for	he <b>2: Ç</b> u	mulative	explaine	d variand	ce of pr	incipal
<b>SZU 32</b> <b>SZU 3</b> <b>SZU 3</b> <b>SZU 3</b>	20 <b>Brooks line to signification</b> 20 320			bitime abregoaints at	etquasinstas Elseptaires	e ifilage pairs21	he <sup>2</sup> CC	mponen	ts of DIN	IO featur	es.	





- Figure 3 unsupervised cluster probe: STE outperforms PCA and Random Projection dim. reduction baselines
- Reason 1: non-linear projection forms distinct clusters (see performance at unreduced dimension  $D_{ViT}$ )
- Reason 2: while there is less information content for lower dimensions, lower dimensions favor k-Means convergend



	5. Conclusion	References
is a it can sion by nance	<ul> <li>Vanilla DINO is highly performant</li> <li>STEGO is a semantics-preserving dimensionality reduction technique, outperforming PCA and RP baselines</li> </ul>	[1] Hamilton et al. "Unsupervised semantic segmentation by distilling feature correspondences", ICLR 22.
EGO (RP) more	<ul> <li>Future Work 1: STEGO's segmentation head can adapt to new data distribution after training using a contrastive loss. Can the segmentation head be trained using the simpler DINO loss, too?</li> </ul>	[2] Caron et al. "Emerging properties in self- supervised vision transformers", ICCV 21.
on ce	<ul> <li>Future Work 2: What is the impact of using clustering algorithms specifically designed for high-dimensional embedding spaces?</li> </ul>	