Application of machine learning to automated image analysis

Robert K. Cowen¹, Moritz S. Schmid¹, Christian Briseño-Avena^{1,2}, and Christopher Sullivan³

- ¹ Hatfield Marine Science Center, Oregon State University, Newport, OR
- ² Department of Environmental and Ocean Sciences, University of San Diego, San Diego, CA
- ³ Center for Genomic Research and Biocomputing, Oregon State University, Corvallis, OR

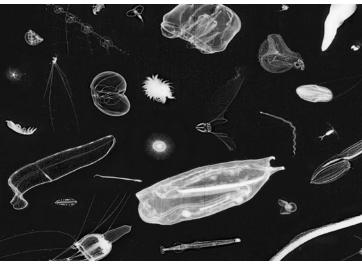


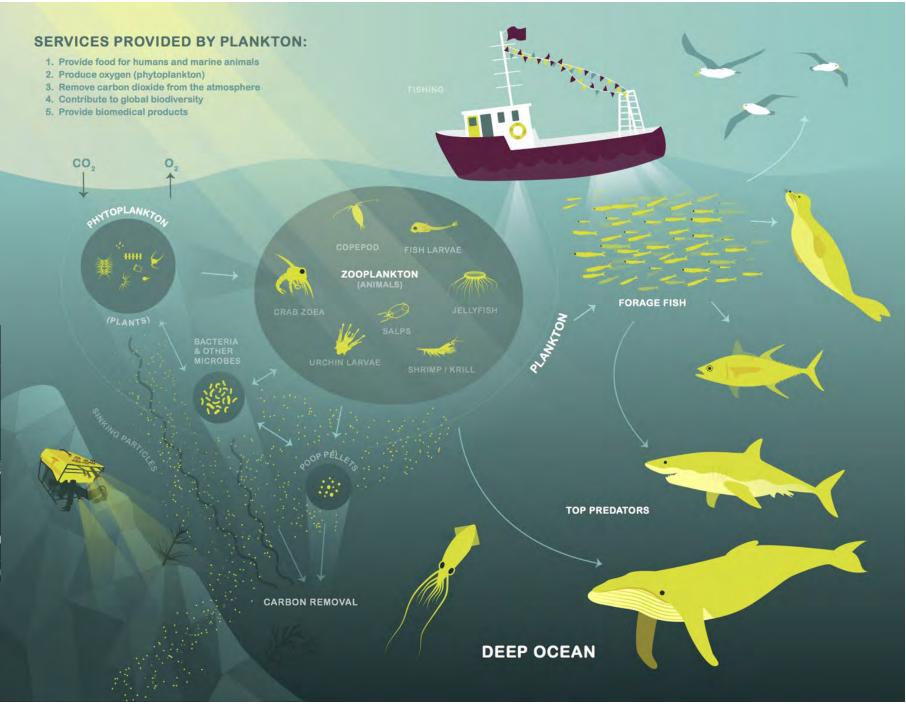






Plankton: How do we sample it?





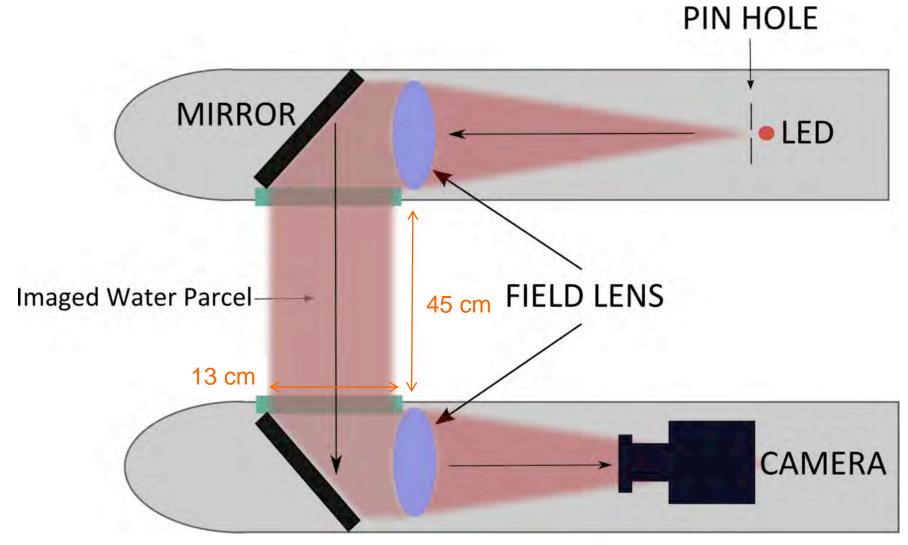
In-situ Ichthyoplankton Imaging System (ISIIS)



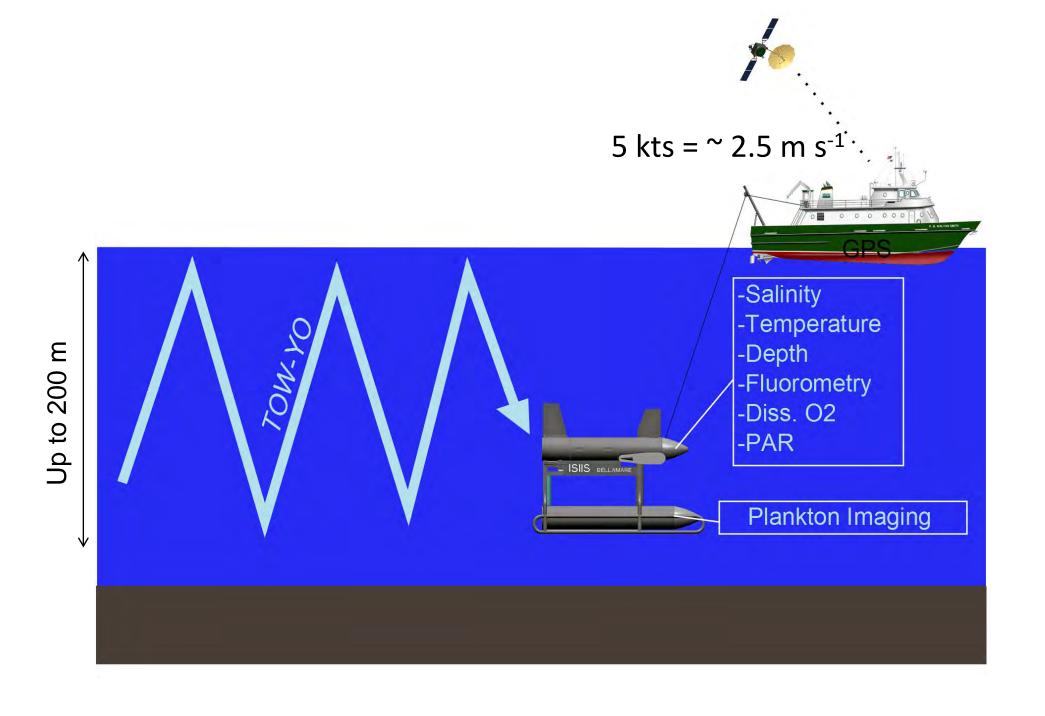


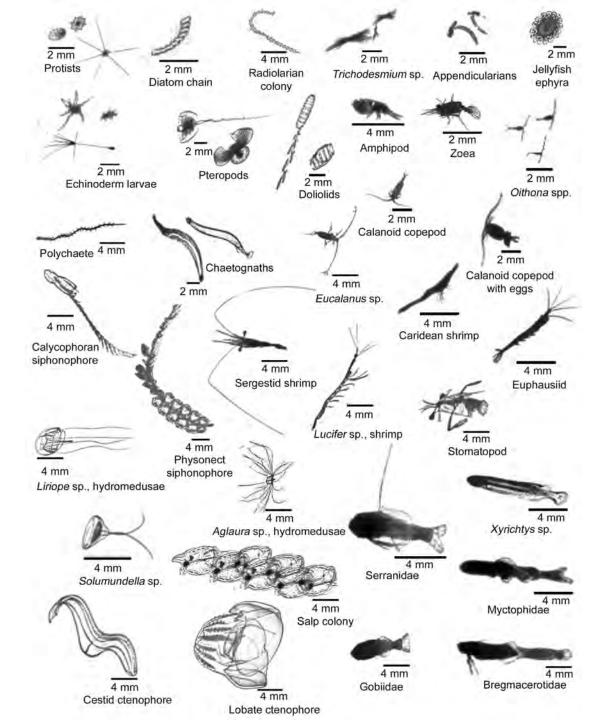




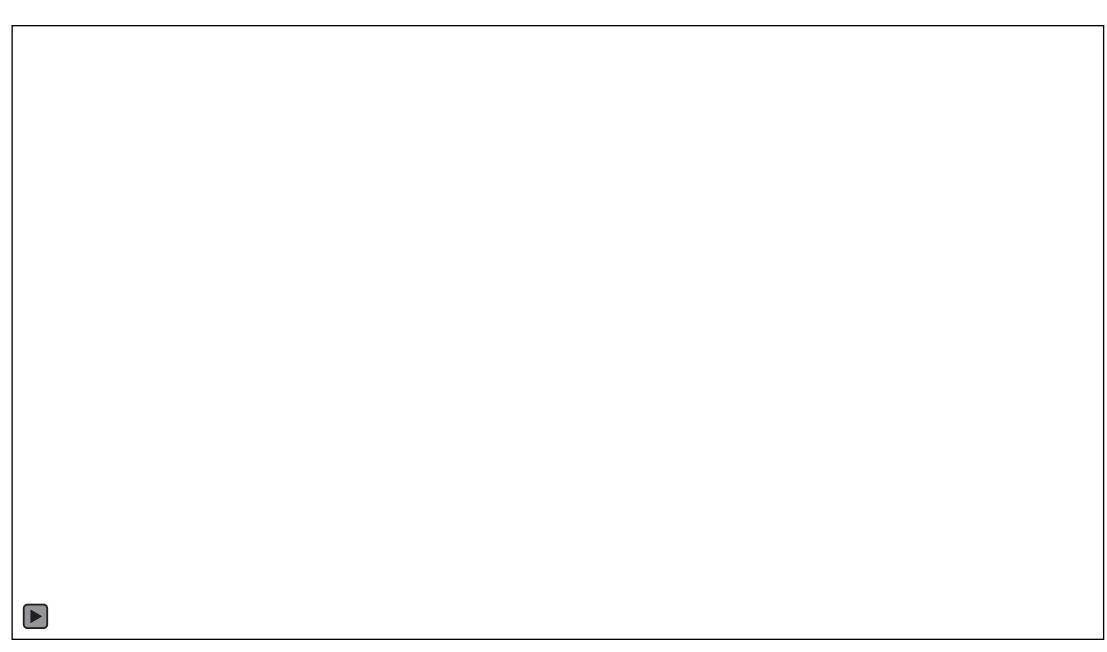


- 13 cm x 13-cm field of view
- 45 cm depth of field
- 2048 pixel line-scan camera
- 66 µm pixel resolution
- 17 frames s⁻¹
- 180 L s⁻¹ imaged





What does ISIIS see?

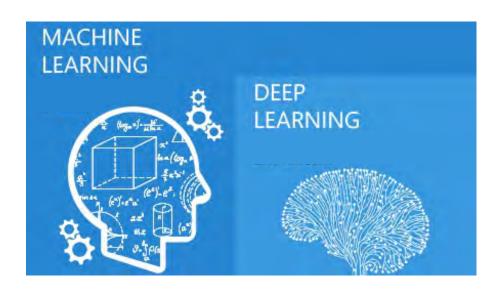


Now imagine that recording going on for several hours

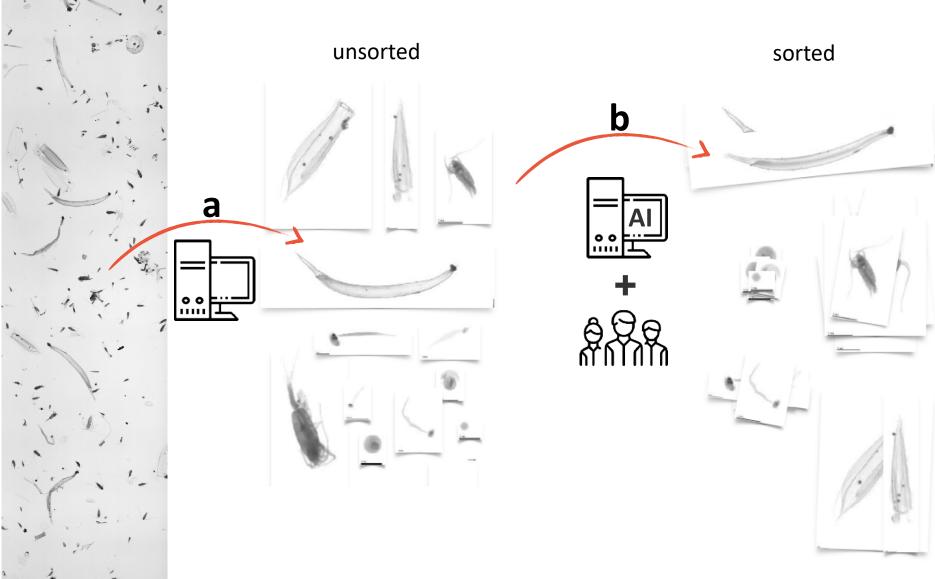


Hundreds of millions to billions of single organisms imaged per cruise





Imaging pipeline - overview

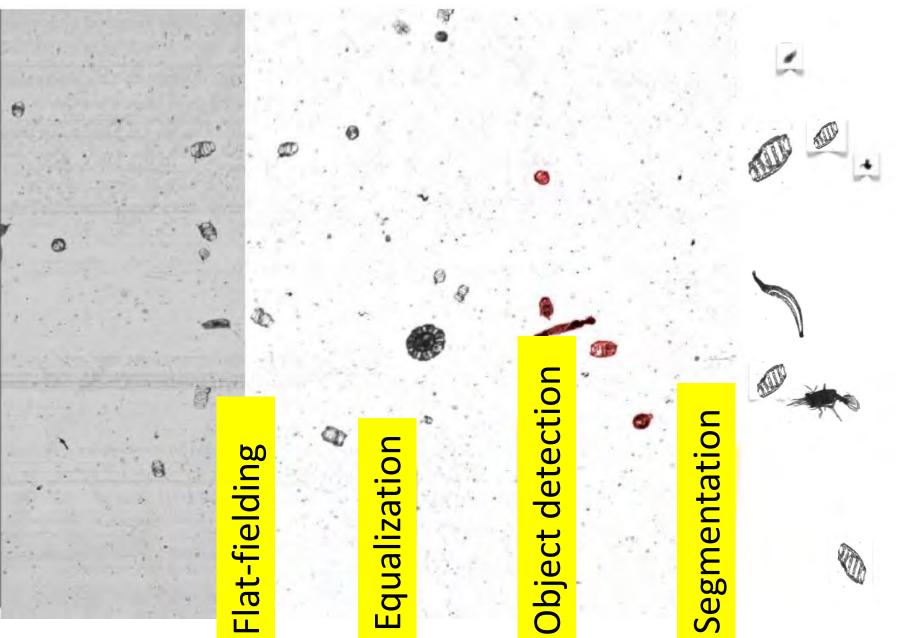


a = segmentation

b = training + classification

vided, M	exies, his citized from	oldest time s	likal itre sideal wa	m pilmi archiv	dimi erelden pener i	dies, mois	use, one by the near	eiden.
1900180103 NO. 1. 1	\$2.306 T.316000067	50180108	120003 velikimel	Lattic Jacket	winter author Signal port	\$0190000	\$4800 1/2012(10) No. 1 1/20	host
1927100:409 bit 7.2	N2.505 7.31506066*	00180-03	120000 reliated	Corinea Destrop	tompolities shrit	02130004	19-20" w20150100 tot 1 2 lbs	India
122 (B) 103 B) F3	12.500 T.3(200066*	20100103	(2000) velianed	Costrae Descon	percudate-vir II	22120004	194344 N2015C1C) to 1 3/00	Merta
PARTERINA PR 1 A.	AT NO. 3 MINORAL	20100000	12000 amones	Contract Descon	negrosigos-arriv	2015000	THE RESERVE THE LARGE	DATE
A.L.M. Extractory	DI OR TATMONRA!	SETHIOUSE.	120331 altisated	Total de Canadas	Nonrection and 1	THERE	NAMES ASSESSED OF TWO	District
DESCRIPTION OF THE	FI SUS T 3150000007	20180103	120007 Garage	Codena Basens	tomos@de-ih-l-	THE STATE OF	HARFY NOTIFIED BY I Find	E,no
ru20100103 tot 1.7	40 300 1/3400094s	20100103	120000 williams	Doinee Descou	Service@clo-shr li	221 30000	184400 (000100100 tot 1 7 kg	derr.
Q\$2180109 84 T 6	12,360 1,31000me.	20180103	120003 miliated	Conser Person	Arm codition with A	2213000A	19-218 420750163 to 1 Blog	bedy
PART HE HAVE BEEN A D.	E2.000 1.31000000*		(2000) williams	Contras Description	annodate-in /	22130604	142701 1020157153 to 1 Blog	
	47.700 1.31000000.	20180103	1,0000 Amounts	Later Andrea	arrogroups /	23150037		dert.
\$276000 M.J.10							18.75m g2016/10_00_1_12.pg	
egictatives for 1 11	PRINCIPAL PROPERTY.	20,180704	120000 vinami4	Solone Disnise	annongou un'x	3318004		
· 11.19 · 10.00 · 10.00 · 10.00	REST LANGUAGE	20160105	120003 veliamos	Correc Doorce	domosilicio vir.h	201 8000	196810 1/2017/17/20 to 1 12/00	
1422400-108_tot_5_10	80,300 7,34500000°	20100103	120000 ellished	Corinee Deenge	semus@do-whrfi	03130000	19-915 (20100100 to 1 12/05	
PARTHOMOGRAPHICS	50,500 3.305000ms*	30100105	120000 will bited	District Beauty	January Street S	33110003	180703 10350105_64_6_164_4	
vyeneous early or	25 AND T NEWSONST	20180108	120000 well-being	Contour Description	hatera@riscale li	131900 la.	HARRY PROPERTY IN THE PARTY	
7320 HONO BY 1 16	RISSET PROCESSO	20140-09	120001 released	Contras Descripti	description at h	30130909	162-4" \$2016165_ER_1_1EEE	-6675
CLT DOLEN WITHOUT SAFE	15.396 3.310000001	25,180708	120000 vmarsid	Largest Associate	probor: partial granul born	3/21/80/6/27	HIN GROWING M. 1.17/99	hydo
NUMBER OF STREET	BLOW TAMBURET	20190108	120003 viname	Coriner Doorse	tore-middleto-vitr.ti	2018/00/7	From NEW SERVICE NA. 1 19,000	1998
rp229800408 tot 6 45	40,300 T.34500006*	20180103	120003 willaind	Dorinee Deenga	Searca (Boto-ehr.)r	22130006	152500 1/20100100 to 1 12/40	dest
1975 TO NO. NO. 1, 20	\$5.300 \$.340000m8*	20180103	120003 pillate-1	Down Descar	Service State and Co.	23000007	MINE WOMENED IN LINE	Dist
synthesia_64_1,21	JOANN TAINMET	20180.00	120000 with Street	Concern Director	Senso milliprosculte fo	23/5000a	Team's STATE OF LOTING	report.
P202100100 00 1 25	42.565 T319666667	901601001	120000 religioned	Contras Deprox	description-in h	10190314	150754 900000103 pp. 1 20 pp.	in the
1982 187 197 MILES	GUID LUDGERES!	25/1807/08	120000 (marted)	Connei Desirbs	poercoappes -em h	221606.0	THE TO SQUARE UP 1 24/00	-049%
PURSUE NO DE DEPUN	D.CO. Laterage	20100100	120003 (0.6664)	Connec Description	Bronneddrin (IPA)	2016007	TOTAL SECTION OF A STATE	
ru22100100 tot 1.00	N. 300 3.30000005*	20150103	120000 utilated	Horizon Deserve	Searcealliste air b	22130000	HINES TODOS SONDO TEX. 1 25 kgs	dert
120 1 St 1 20100103 pt	NO 305 3 545000000*	55180-00	120000 withhold	Lastin Initiat	eliger, indist@cost.com	2313000a	162353 1/20130100 (ut 1 20 to	
OFFICE AND PERSONS	STAM TRIMMART	dictac do:	120000 pinned	Research Pilotopic	Street Charolinians	20190719	NORT SHIPTING LIFTING	Date
OWNERS OF LAKE	AT RIPL'S STREET	Sentences	120301 altifolise	Driema Basses	him regiondr t	THERE	STORY CONTROLS SELLINGS	
PERSONAL PROPERTY.	13.515 T.318009947	20190100	120000 velidend	Corinee Dearce	Searce(Date-Art)	92180000	19-319 1/2018/10 lbt 1 28 kg	
400180400 NV 1.50	13 916 1.310000067	20190103	120000 polidenia	Conces Donnes	Assertant Ballet yields	90190000	140701 subbracket as 1 80 km	
VARTING NO. No. 1 51	\$2.300 3.30000984*	30160109	120000 selikind	Contract Description	intercollisionals 5	22190907	troute sportered by 1 Man.	
F262 180 103, 101 1 22	EL300 1310000007	20180103	(2000) addition	Driften Descrip	awordow-in/	200802	18 344 1/20150 ED to 1 32 cm	
DALMARIA BATTO	ED THE I PROGRAM.	20780700	1/9900 amount	Contrae Descore	menonium ar t	2018/00W	"0014 NORMANO ET 1 2025	
SALES AND THE PARTY OF	To con I streame.	Setterner	170301 4604641	Drinne Hasens	Manager and E	Allent.	transi gamento ne i save	
THE PART OF SALES	ELSE THREESE	20150-03	120000 velidebe	Corine Disprop.	torrosilicio «ir.)	2018007	170407 N209 E1100 Nt 1 20 pp	
CASSAGE 408 FOR A DC	45 305 3 34500000°	20180103	120000 validated	Corne Denos	Servingers -W.)	12130007	176866 og0930100 mz i bilios	
14E7100109 NA 1 09	45,365 3,312000004	\$E180108	120003 millulated	Do'n my Desiron	ine unlikitorek li	13000103	18040" NEOTICE IN 1 OF NO.	
ATEL BOOKS THE T. SE	27 SAN T SAMMAKET	27 (40 -0%)	120707 militared	Distinue Reserves	Special Control of the Control of th	13190006	THE STREET, IN LINES	
decade of the Property	N1595 T.E0000447	20190-09	120000 renormed	Corinea Deeros	sercegore-m.h	22190W	HAVE STATE OF 1 34 PE	
Section and section	PRINCIPAL PROPERTY	25180108	120002 variable4	College Design	Best-ologon: JP-X	3316000	152791 SECTION OF LANSING	
THE PART OF THE PART	STATE LANGUAGE.	20100108	120000 velidated	Lockita Anobest	sloter, settlel fignel ter	2018000	CONTRACTOR OF LAYER	
1922100100 tot 5 42	40 300 T 34500000°	20180103	120000 reliated	Coinse Descar	ann collaborate 5	22130007	170255 1/2010210 to 1,42,60	
WITH AND PARTY OF	50.500 3.5000000*	20180105	120003 selided	Contract Description	Jenson Quincally A	22130009	151619 1/2001/2105_64_1_65/46	
NUMBER OF STREET	27 NAT 1 NIMAGET	37 (86 (9)	120000 will benefit	District Dissorts	harvelly to air b	POINTS:	OF MAY PROPERTY OF 1 STATE	
rgazinoros en 1,46.	KL565 TJH566666Y	20140108	120000 religioned	Contras Degross	derivation-in h	30430907	HOME SOURCE OF LANGE	
MATERIAL STREET	SUR JURNAMET	25/18/21/04	120000 (600004	Donné Disnos	poenusggtav virt h	STREET	191400 19009411C; 00 1 44.98	9905
PARTITION FOR SALE A.	D.695 1.35866867	25,180,103	120000 4006068	Colema Danam	processibles with the	20190004	TRUSH NEW WILLIAM I AT US	766
24.1 Fel 800-000-024	40,305 3.36500006°	20190103	120000 williamed	Dorines Desiras	asarca@clu-shr3*	2200007	180700 1000130100 tot 1 45 kg	
PARTITIONS NAMED AND	\$0.305 X 35000000*	50180108	120000 1140-4	Cowne Downe	Administration of the Company	23000007	THEIR SPHIRTS IS LANG.	Design
AND REPORT OF THE PERSON	COMMITTEE THE CO	30166109	120005 editions	Serveria Filmor	energialming twell	20/20010	HOME WASHINGTON OF LAND	Date
rg20480403, eq. 1.51	A1.505 1.310000007	20160-00	150000 wildsawd.	Contras Deserves	nerogion-int	10430000	issasa waxisticic) on i bigg	
PROTECTION 101 1 50	NAMES ANDRESSES	201907-09	12000 (mare)	DONNA DAGAGE	2001/06/000H - (PF)	301160009	THE OF RESPUENCE OR 1 SAME	
PART - 10 MILES	IS SIG TATRICIDAY	20180103	120000 polisima	Gorney Dogram	Scorrandibiles all-A	20130004	100010 100010101 to 1 63 cm	

Segmentation

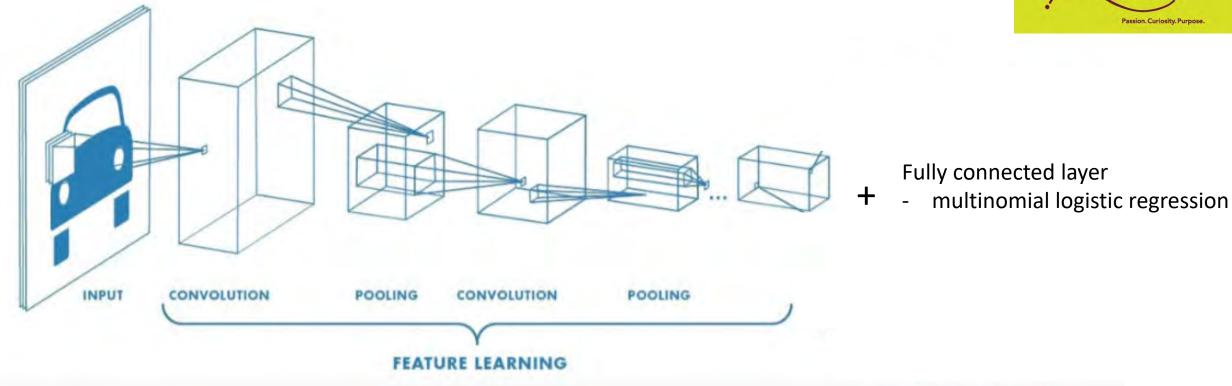


- Unsupervised k-harmonic means clustering
- Bounding box;>= 250 pixelsare retained
- CPU machines (100+ cores)

Convolutional neural networks (CNNs)

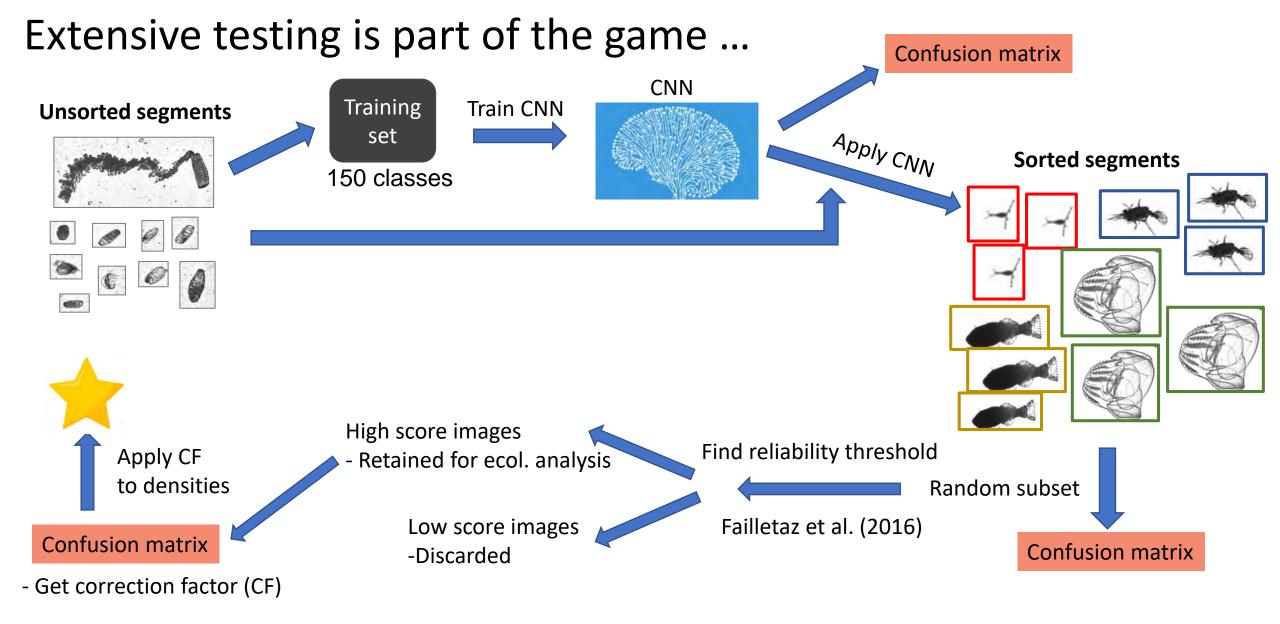






- Convolution (13x) = detect local combinations of features in images
- Pooling (12x) = merging similar features into one

SparseConvNet with Fractional Max-Pooling (Graham 2015)



Current learning set: 161 classes

5.5h transect = 115M segments = 55h of processing (8 parallel jobs on Nvidia P100 GPUs)

Efforts to make the pipeline more broadly available to the imaging community

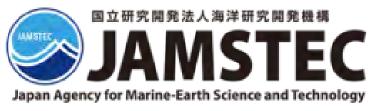
- www.PIC funded by Belmont Forum
- Streamline steps to library generation
- Optimize CNNs for use across imaging systems (UVP, ISIIS....); transfer learning
- Make pipeline available to others









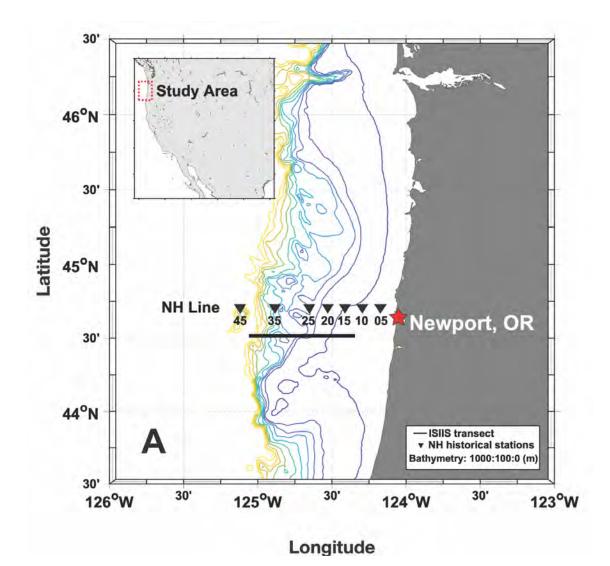


Three-dimensional cross-shelf zooplankton distributions off the Central Oregon Coast



NOAA ship Bell M. Shimada

- Newport Hydrographic Line, sampled since 1961
- June 2016



 What could be gained by adding an underwater plankton imager to more traditional deployments?

