Deep Mixture of Diverse Experts for Large-Scale Visual Recognition

Jianping Fan Dept. of Computer Science UNC-Charlotte, NC, USA

Outlines of Presentation

Research Motivation

- Deep Mixture of Diverse Experts
- Ontology-driven Task Group Generation
- Deep Multi-Task Learning, Deep Boosting & Deep Collaborative Learning

Large-scale visual recognition

Conclusions & Future Work



1. Research Motivation

Many applications rely on large-scale visual recognition





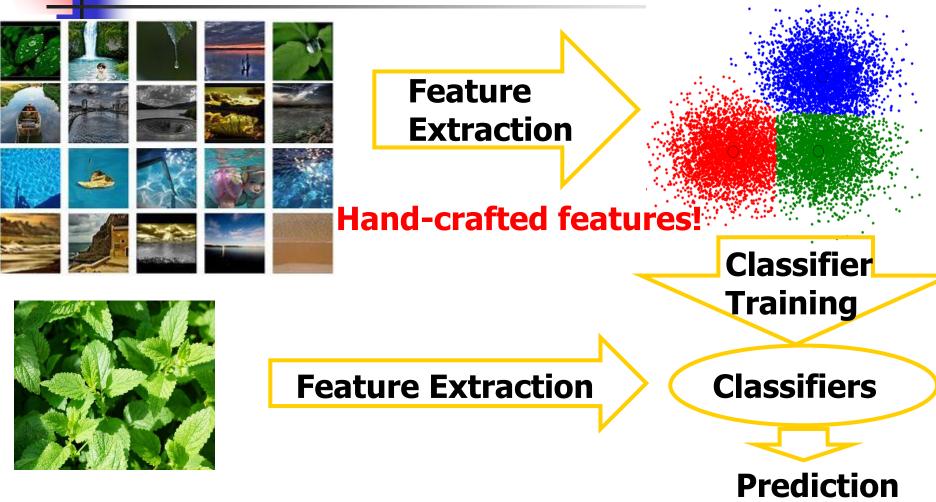
document electronify

surveillance

elder care

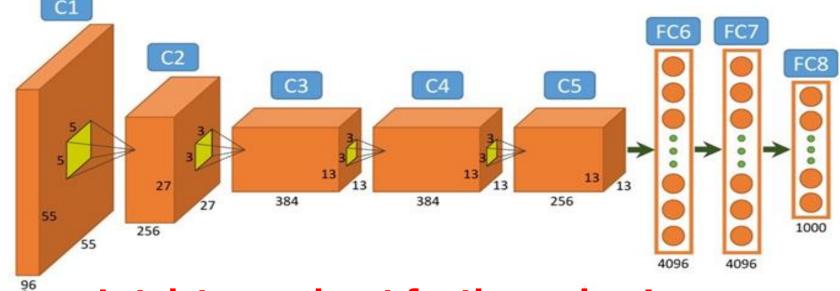
Traditional Solutions

Separate processes for feature learning & classifier training



Deep Learning Approach

Joint process for feature learning & classifier training



Let data speak out for themselves!

SGD for back-propagation

Flat Softmax

- Problems of Flat Softmax
 - Inter-class visual correlations are completely ignored!
 - Differences on their learning complexities are completely ignored!

 Back-propagation may pay more attentions on hard object classes but it may easily achieve higher accuracy rates on easy object classes!

Traditional Deep Networks

- Problems of Traditional Deep Networks
 Network structures
 & node weights
- They are optimized for recognizing 1,000 object classes or less than 1,000 classes!
- They train one joint network for the hard and easy object classes even they have significantly different learning complexities!

Plant Species: more than 200K Taobao Product Categories: over 100K ImageNet10K: over 10K



Some well-designed traditional deep networks

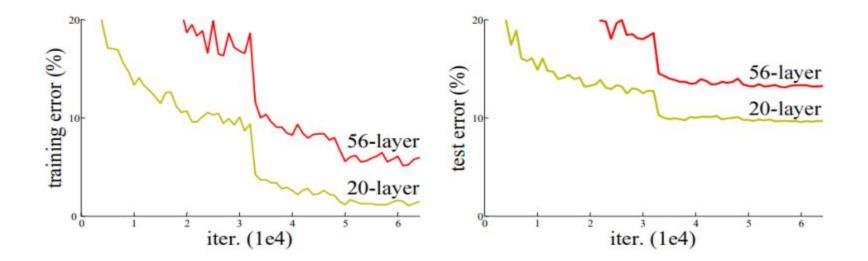


Recognizing 10,000 categories or even larger

How to Configure Huge Deep Networks?

Potential Solutions for Network Extension

 More layers & more units on each layer, ---too expensive for trials & errors, & may not be doable for most academic researchers



How to Configure Huge Deep Networks?

Potential Solutions for Network Extension

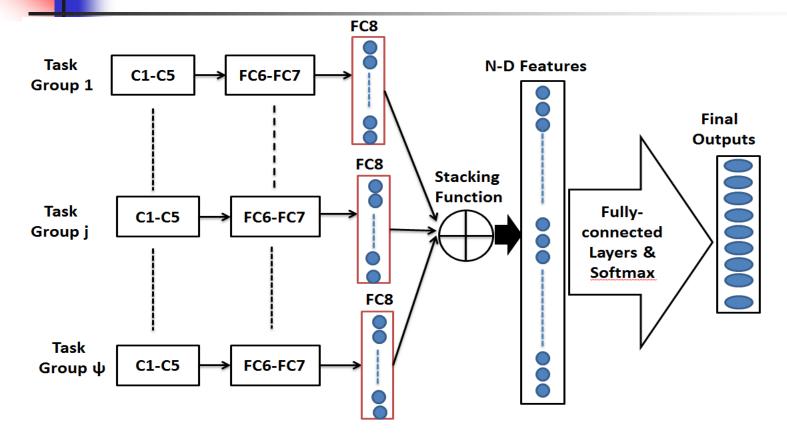
Deep mixture?

---they require same task space

Transfer learning?

----they can transfer the common knowledge from larger task space to smaller ones

Simple Mixture & Random Task Group Generation



- **Problems:** Some bench NBA players may become MVP in CBA
 - L. James vs. M. Jordan

Simple Mixture & Random Task Group Generation

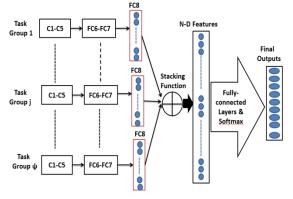
Global Optimum

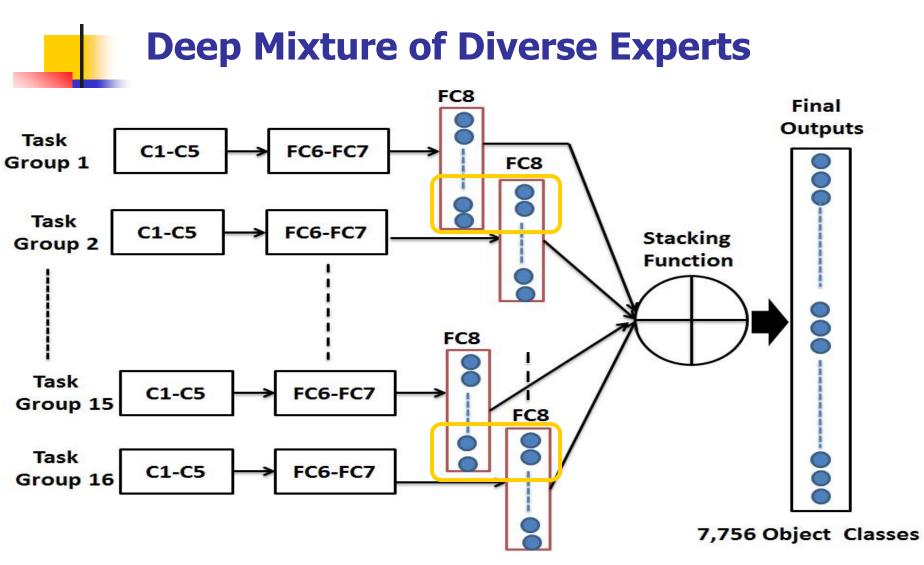
--- The gradients of objective function are not uniform for all the classes in the same task group!

Task Assignment for Group Generation

- ---Different classes may have significant differences on their learning complexities!
- Prediction Comparability

---L. James vs. M. Jordan



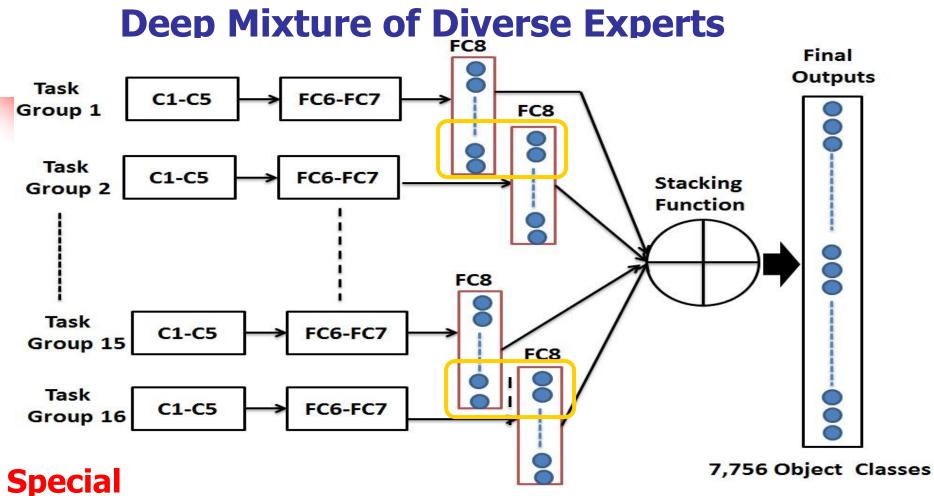


Deep Mixture of Diverse Experts

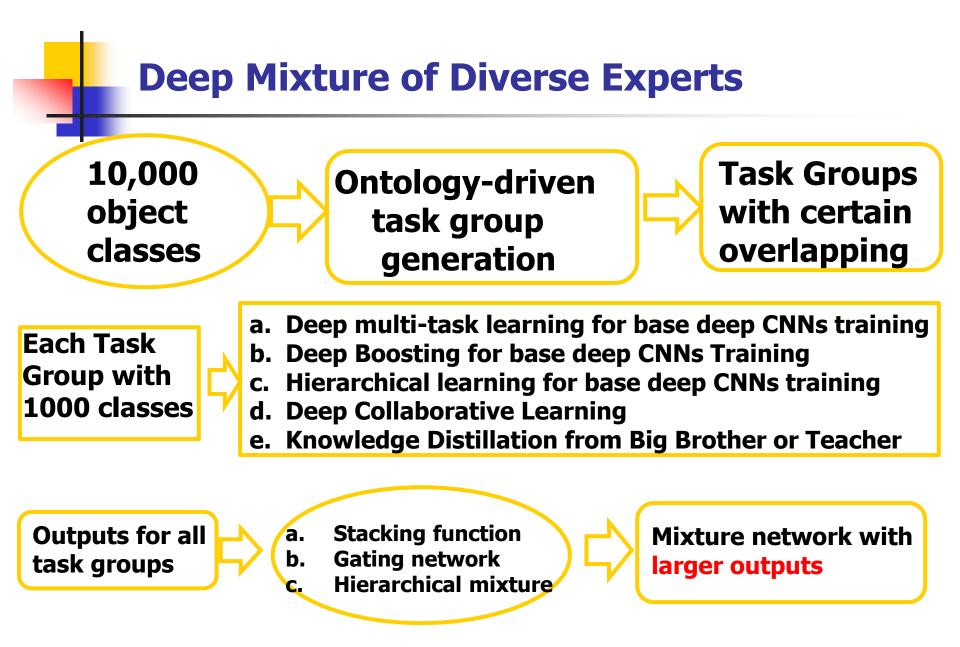
- Potential Problems:
- They are not learned jointly, thus their predictions may not be comparable directly!
- For each test sample, all these base deep CNNs will provide their individual predictions, which one is more believable?
- Object classes may have different learning complexities!

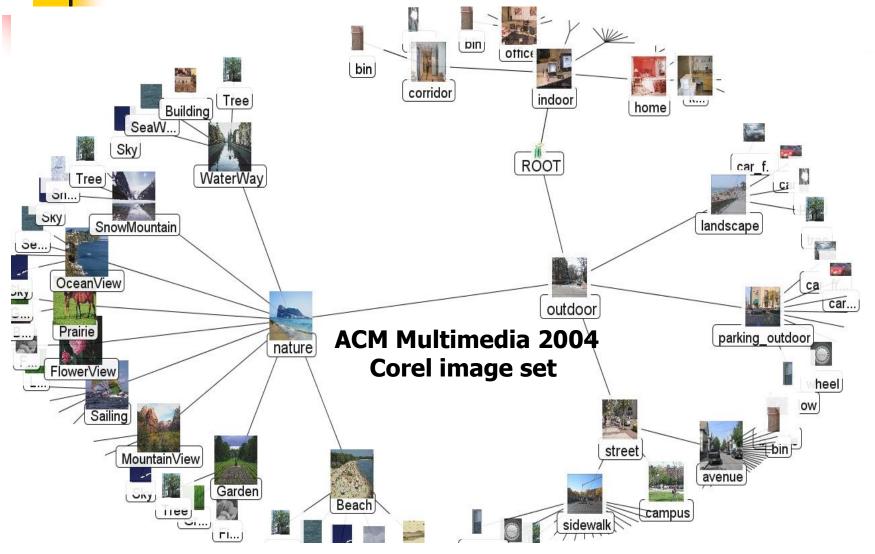
Deep Mixture of Diverse Experts

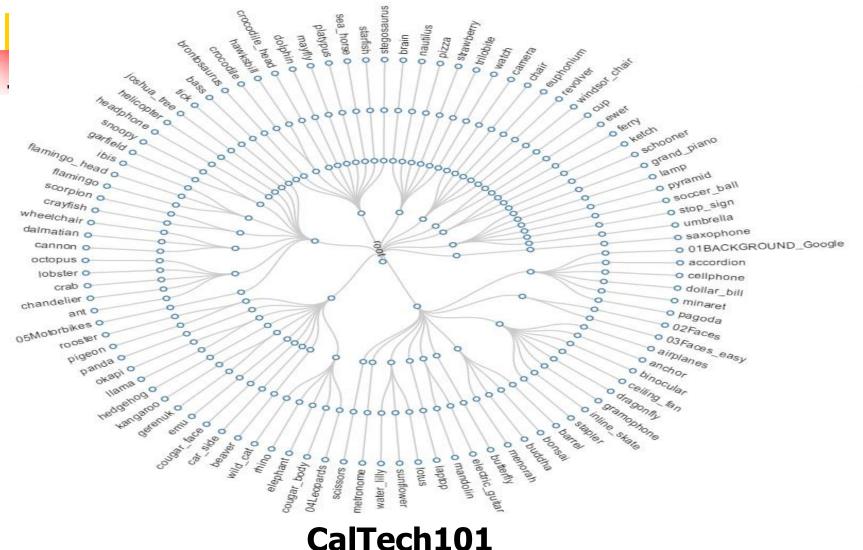
- Wish List:
 - The object classes with similar learning complexities are assigned into the same task group!
 - The predictions from multiple base deep CNNs could be comparable at certain level!
 - The prediction conflicts among different groups are predictable or identifiable!

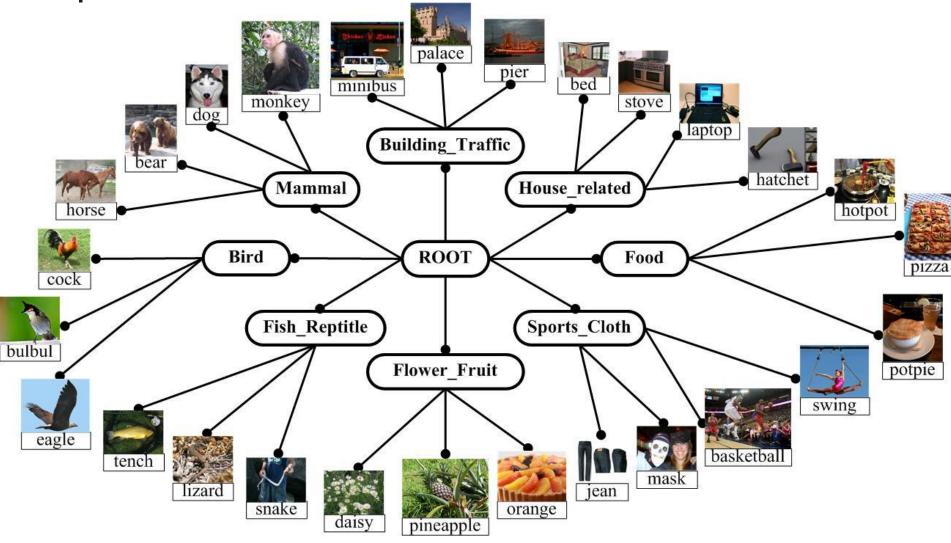


- . Design
- Object classes with similar learning complexities are assigned into the same task group;
- Task overlapping to enable inter-group communication;
- Special class of ``not-in-group" to enhance comparability;
- -----



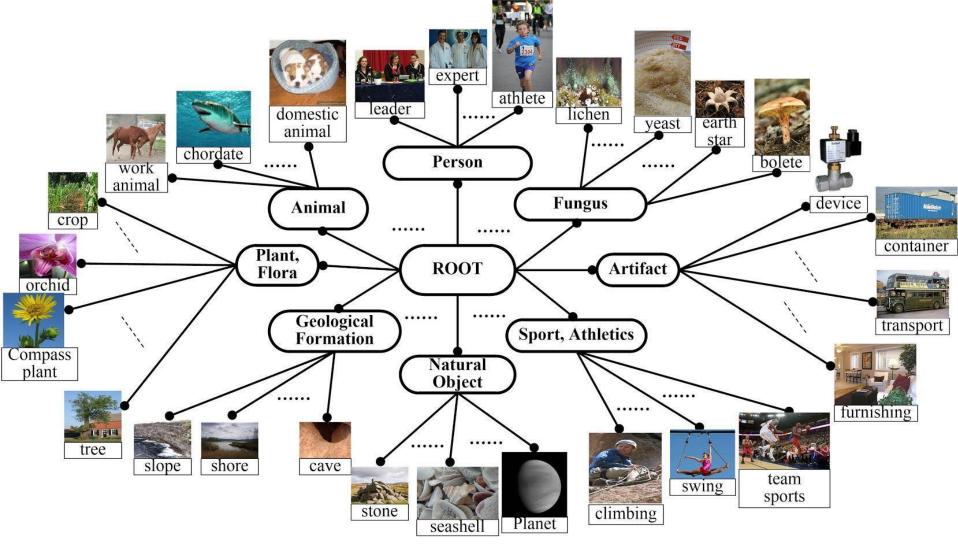




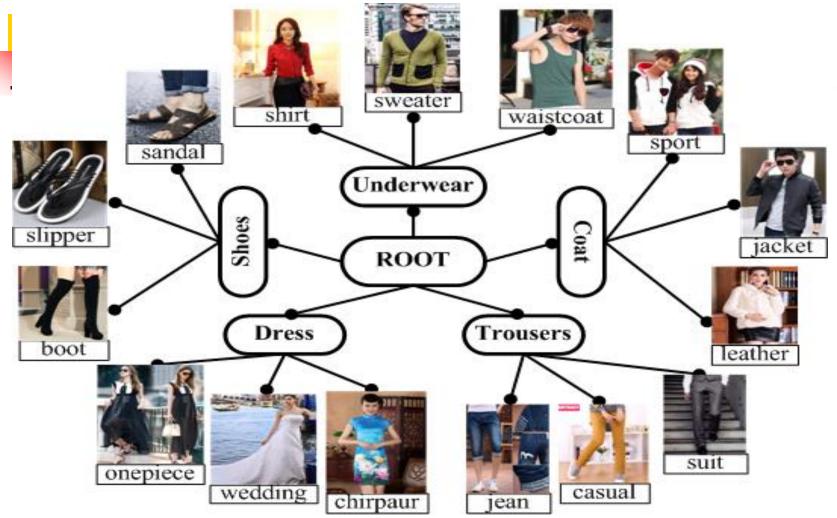


Two-Layer Ontology for ImageNet1K

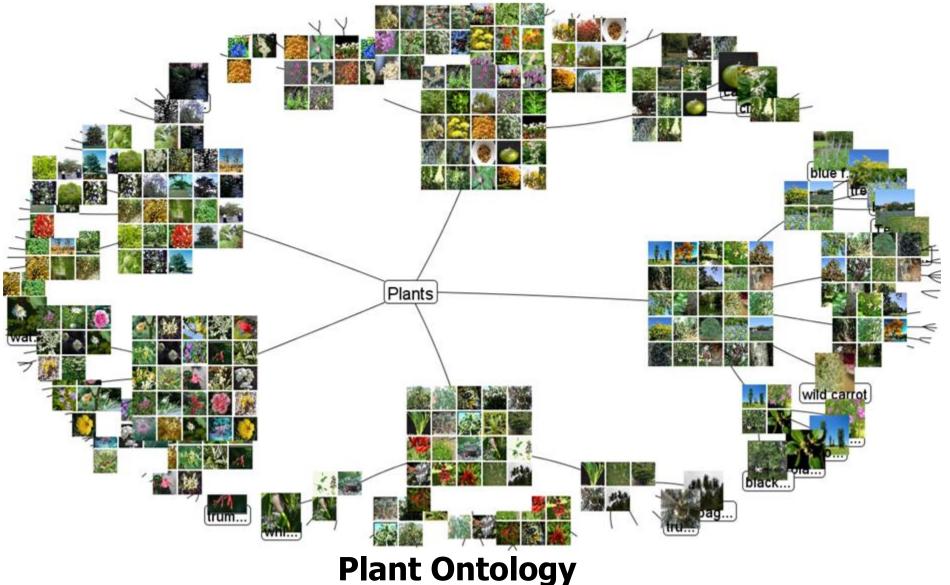
											ImageNet1000													
Exercise related (device)	Daily tools	Kitchen and food (tools)	Package	Paper-related (books, characters, pictures)	Cosmetics	Hats and masks	Cloth (clothes, shoes, others)	Light and fire	Furniture	Shops (various goods)	Electronic device (phone, computer, device)	Instruments	Weapons	Traffic tools	Machines	Structure-related (building)	Insects	Reptiles	Birds	Fruits	Marine creatures	Monkey-like animals (monkey, gorilla, ape)	Animals with four legs	Dog-like animals (dog, wolf, fox)
\square	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box	\Box
20 classes	44 classes	72 classes	13 classes	9 classes	9 classes	19 classes	62 classes	10 classes	31 classes	12 classes	53 classes	26 classes	9 classes	79 classes	27 classes	63 classes	41 classes	36 classes	60 classes	39 classes	49 classes	20 classes	66 classes	131 classes

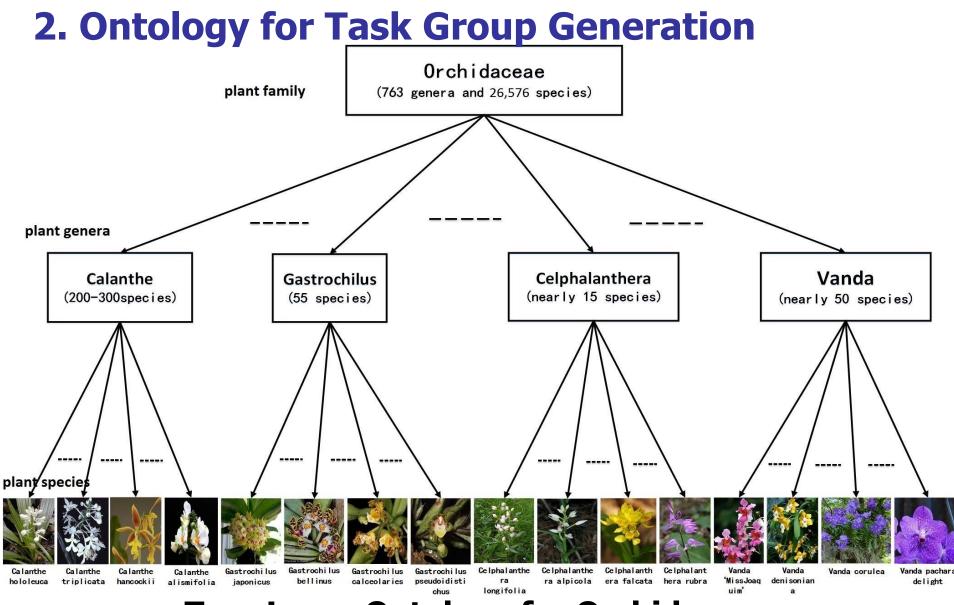


Two-Layer Ontology for ImageNet10K

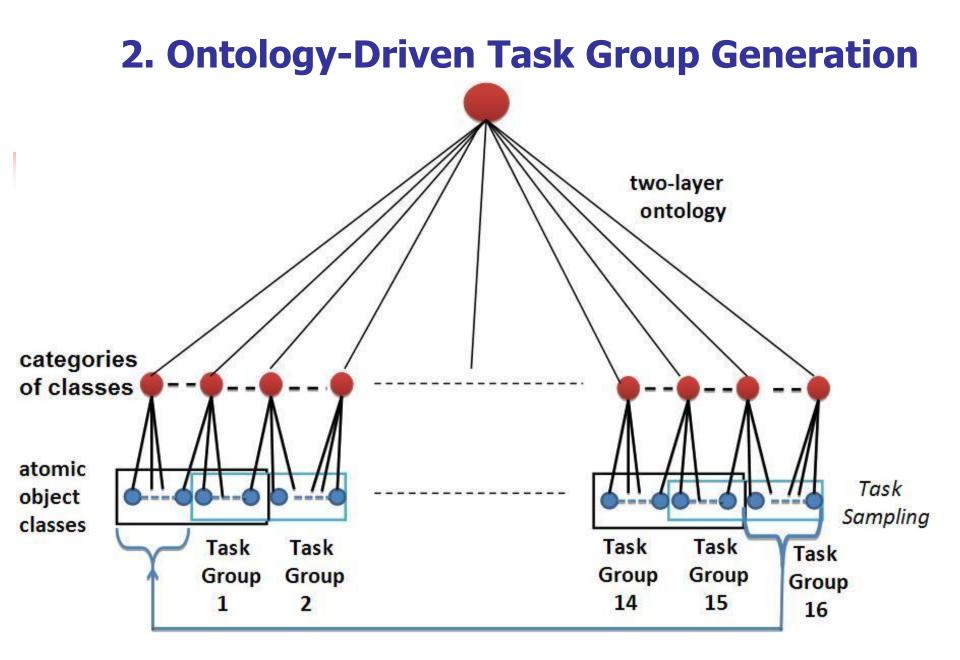


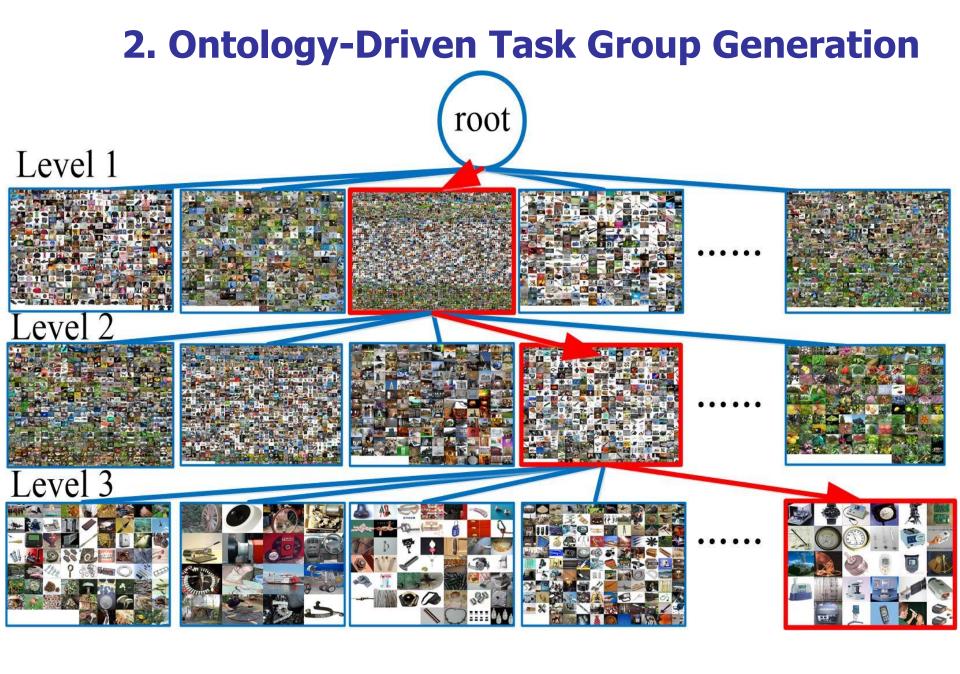
Two-Layer Ontology for Taobao Products



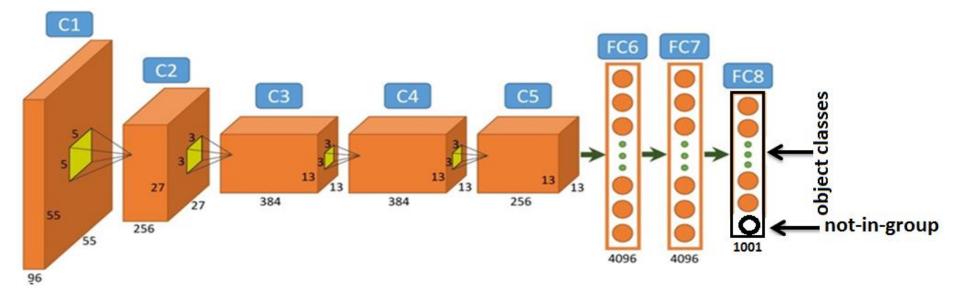


Two-Layer Ontology for Orchidaceae





Design of Base Deep CNNs

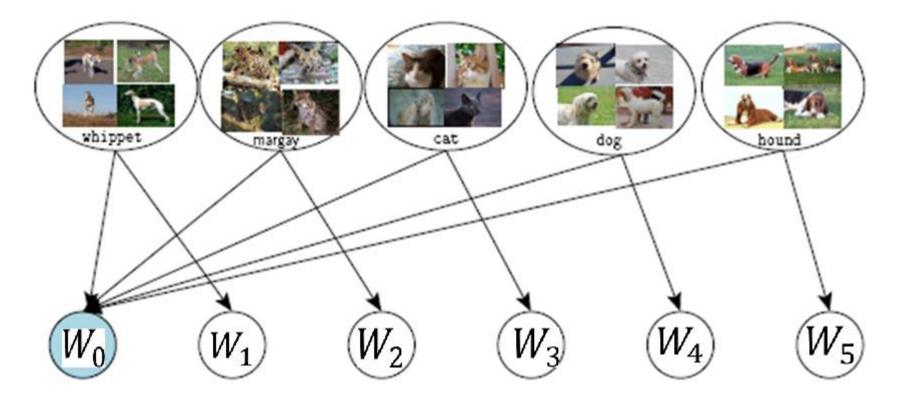


AlexNet, VGG, GoogleNet, ResNet,

MobileNet can be selected for smartphone applications!

- Approaches for Learning Base Deep CNNs
 - Deep Multi-Task Learning
 - Hierarchical Deep Multi-Task Learning
 - Deep Boosting
 - Deep Collaborative Learning
 IEEE TIP 2017, 2018, PAMI 2017, 2018

• Deep Multi-Task Learning: $F_j(\mathbf{x}) = (W_0 + W_j)^T \mathbf{x} + \mathbf{b}$



Deep Multi-Task Learning

$$\min\left\{C\sum_{l=1}^{R}\sum_{j=1}^{B}\xi_{j}^{l}+\delta_{1}Tr\left(WW^{T}\right)+\frac{\delta_{2}}{2}Tr\left(WLW^{T}\right)\right\}$$

subject to:

$$\forall_{l=1}^{R} \forall_{j=1}^{B} : y_{j}^{l} (W_{j}^{T} \cdot x_{j}^{l} + b) \ge 1 - \xi_{j}^{l}, \ \xi_{j}^{l} \ge 0$$

Deep Multi-Task Learning

$$\min\left\{\sum_{j=1}^{B}\sum_{l=1}^{R}\beta_{l}^{j}-\frac{1}{2\delta_{1}}\beta^{T}Y\Re\left(\Re+\frac{\delta_{2}}{\delta_{1}}\Re\left(L\bigotimes I\right)\Re\right)^{-1}\Re Y\beta\right\}$$

subject to:

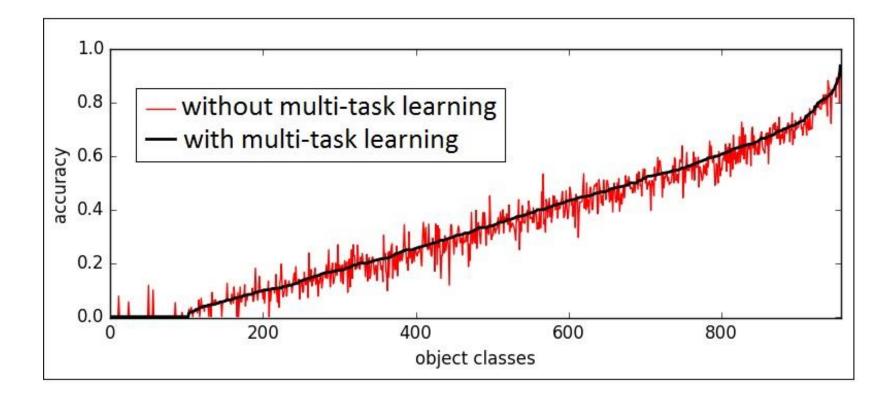
$$\forall_{l=1}^R \forall_{j=1}^B: \quad \sum_{l=1}^R \beta_l^j \cdot y_l^j = 0, \quad 0 \leq \beta_l^j \leq 1$$

3. Learning Base CNNs for Each Task Group Deep Multi-Task Learning $\alpha^* = \frac{1}{2\delta_1} \left(\Re + \frac{\delta_2}{\delta_1} \left(\Re \left(L \bigotimes I \right) \Re \right)^{-1} \Re Y \beta^* \right)$

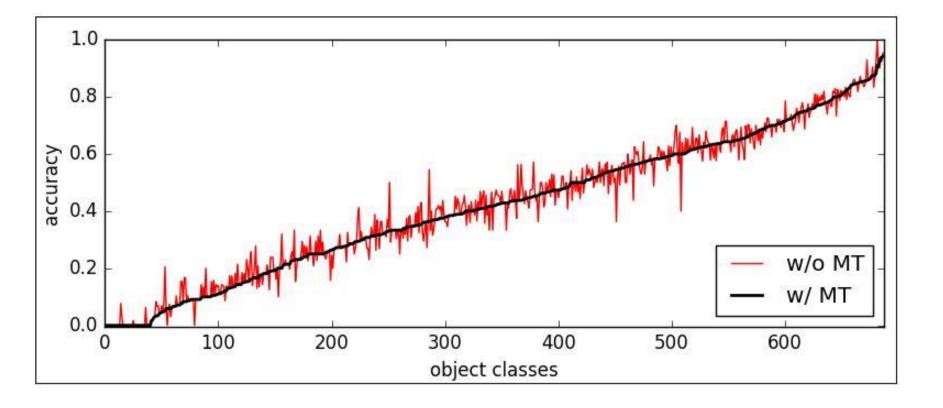
Multi-Task Classifiers at Sibling Leaf Nodes

$$\forall_{j=1}^{B}: f_{c_{j}}^{1}(x) \mid_{F_{c_{j}}^{1}} = \sum_{l=1}^{R} \alpha_{j}^{l*} \kappa(x_{j}^{l}, x) + b_{j}^{*}, \ c_{j} \in c_{h}$$

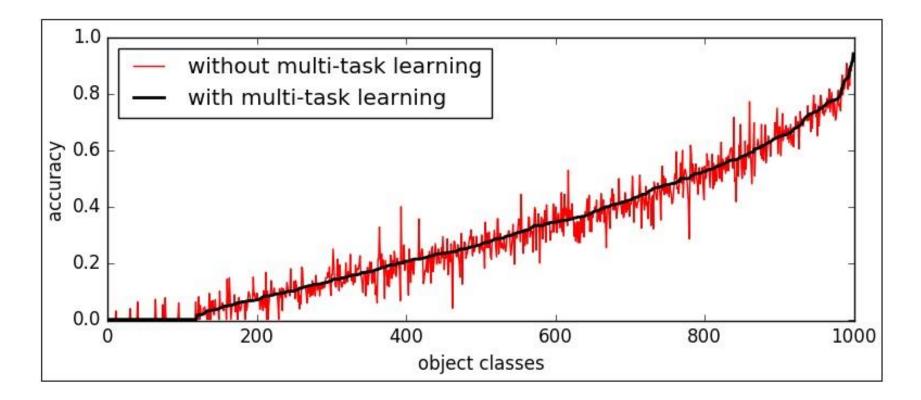
Deep Multi-Task Learning



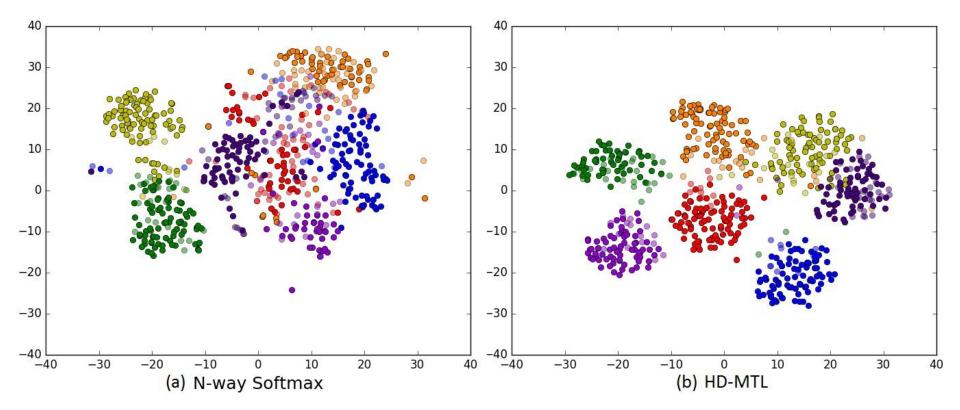
Deep Multi-Task Learning



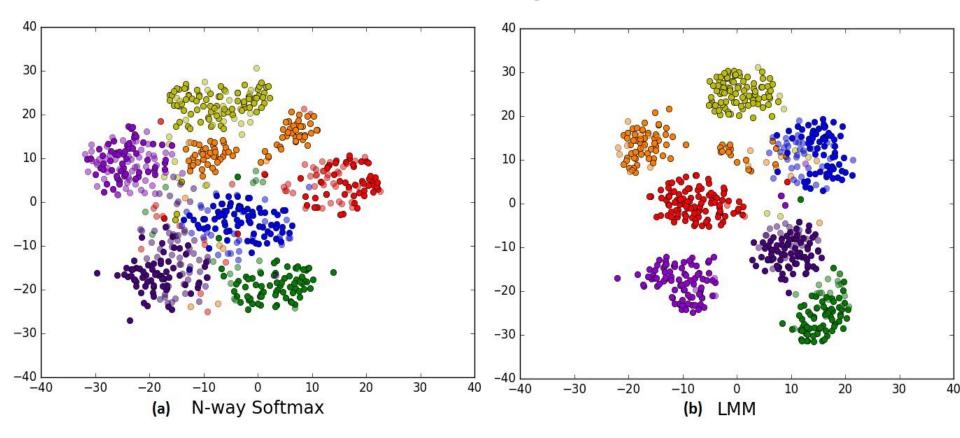
Deep Multi-Task Learning

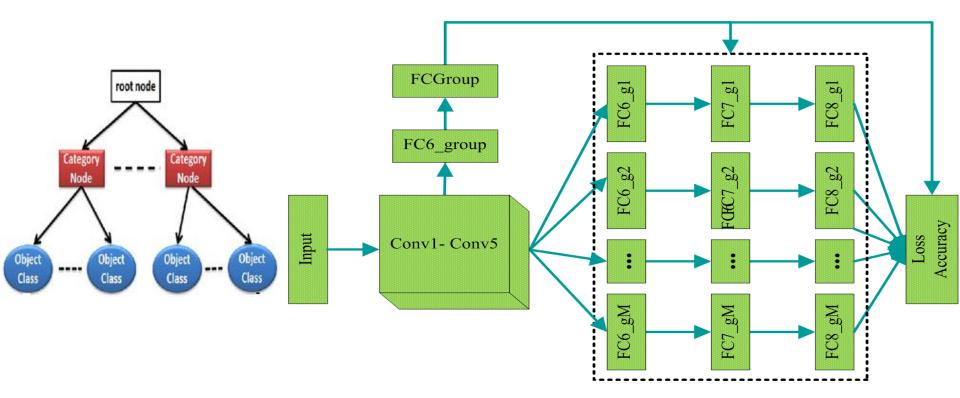


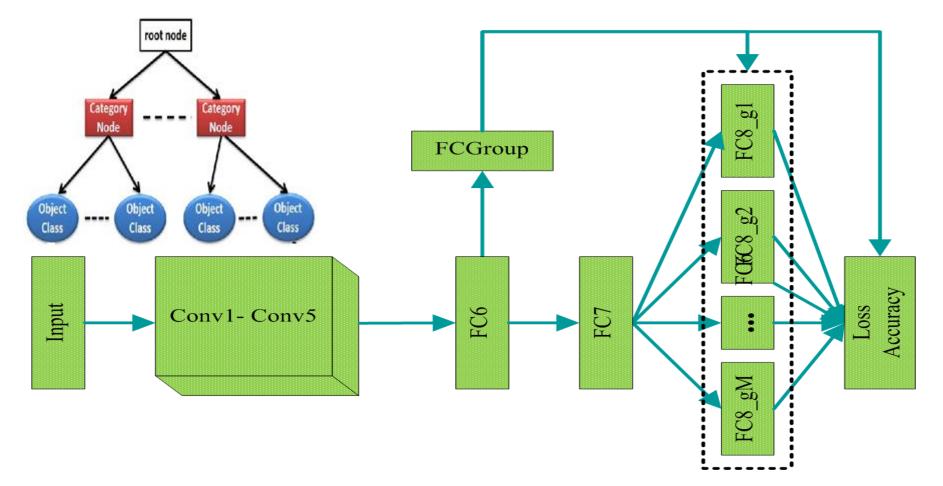
Deep Multi-Task Learning

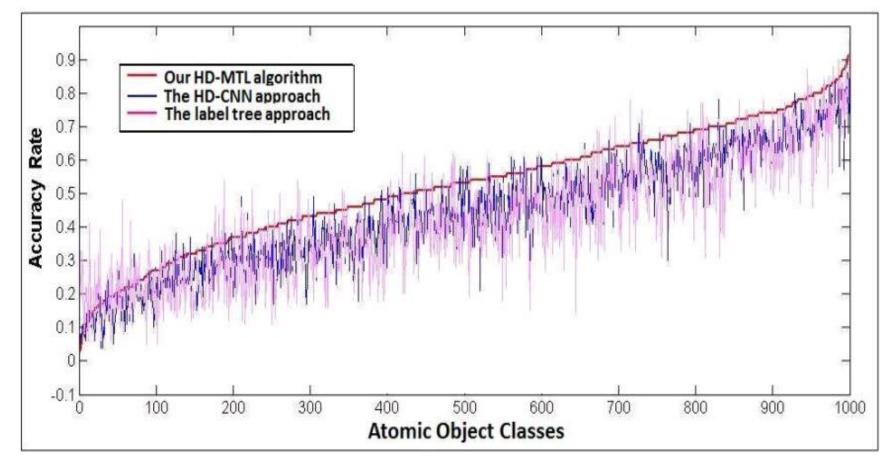


Deep Multi-Task Learning

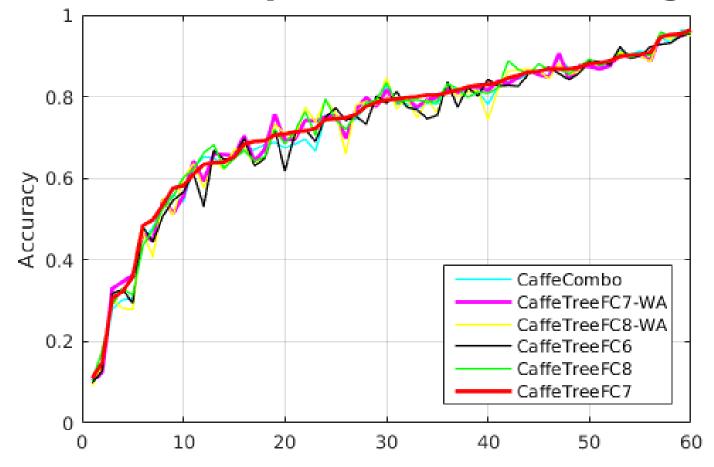








3. Learning Base CNNs for Each Task Group **Hierarchical Deep Multi-Task Learning** 1 0.9 0.8 Accuracy 0.7 0.6 CaffeCombo CaffeTreeFC6-WA CaffeTreeFC7-WA 0.5CaffeTreeFC8-WA CaffeTreeFC6 0.4 CaffeTreeFC8 CaffeTreeFC7 0.3 20 40 60 0 80 100



beaver øtter geal dølphin whaie	vehicles_2 large_latural_outdoor_scenes (sin large_man.made_outdoor_thing aquatic_mammab		aquarium_fish frout shark ray flatfish	reptiles Noven Raag_obntainers Krees		cup plate powl bott e can	vehicles_2 people household_formiture household_electr cal_devices
television Reyboard Imp clarck telephone	inthicks sourcehold furniture arge_containers rood_containers household_exectince_devices	-	forest mountain sea plain cloud	ventriles_2 	- 3	mountain forest Cloud Plain sea	cons artis_mon-made_outdoor_thing tsh artis artisaturaoutdoor_scenes
snail worm iobster spider cro	natural_outdoor_scenes	S-	boy baby man giri woman	arge_annivores_and_herbivor mail_mannals aquatk_mann arge_cannivorr s geople		man woman girl baby boy	nossehold_furniture
palm_tree oak_tree pine_tree willow_tree maple_tree	reatins arge_man-made_outdoor_thing fab arge_matural_outdoor_scenes rese	A	bicycle motorcycle pickup_truck bus train	tool_containers fish arge_mati-made_outdoor_thing household_electrical_devices vehicles_1		rocket lawn_mower streetcar tractor tank	eessie nousehold_electrical_devices vehices_1 herefhan-made_outdoor_thing vehices_2
whale Beal delphin begver otter	arge_natura fait reptiles arge_omnivores_and_herbivor auutic_mammais	X	bottle can plate bowl	mit_and_vegetables household_sectrical_devices household_netebrates household_turn food_containers	-	chair Couch Ded Wardrobe table	lood_containers reptilex rees arge_natural_outdoor_scenes household_furniture
apple weet_peppe mushroom pear orange	household_electrical_devices r housefold_furniture howern food_cor tainers fust_and_veget.	5	tiger leopard bear wolf	small_mainmals people large_omnivorps_and_herbivory medium_mammals large_carnivores	(19)	turtle crocodile lizard dirosaur snake	household_electrical_devices medium_mammais large_camivores large_camivores, and herbivor reptiles

Deep Boosting

- Hard object classes may have higher learning complexities, but easy ones may have lower learning complexities;
- Learning a joint network for both of them may not make sense, e.g., their errors may have significantly different effects on optimizing their joint objective function!

Deep Boosting

Visualizing object classes according to their Learning Complexities

IEEE Trans. on Multimedia, 2012

Visualization of whistle mortarboard oupe, jeweler' opolo tie, bolo, thimble mousetra academic down. **Easy Classes** knee pad boy **Seperability** puck, hockey pu ping-pong ball cassette GD player addle, boat pa broomab, swob, mop binder, ring-bi cassette plaver mard disc, hard mwitch, electri oudspeaker, sp dapten kaptentet cowboy boot shower curtain modem nipple brassiere, bra projector monitor . nematode nemat space bar computes key houtescreen, CRT scr typewriter keyb desktop compute inion soap dispense bikinintaiöebiéank s dypewriter keyt saltshaker npúte elevision tel iovstick envelope hatchet Gardigrake Crashib bathing cap, barriegetic guiterickelhaube rationater hammer beagle dendar basset, basset power drill cleaver, meat c home theater. chain saw, chasleeping bag se dumbermill, saw basenii ho, cohoe, co assault new lver, six-a bloodhound. triceratops file filebookitase candle taper dire screen, f trilobite volleyball platypus, duckbippopota tingray butcher shop, m guinea pig, Baary shieldsbeitkmeil, rin barrow, garde china cabinet, gondola necklace oil filter vacuum, vacuum heetter German short ha shopping car arocerv store. ... Baek wall clock cuirass thatch, thatche plate rack diver windbuildiver volcano andbar, sand ba accordion, pian 🖲 sewing machine smoothing breastplate, amarmot black grouse sundial strich. Struth steel drum turnstile bucket, pail bustard geyser duaona Daona seashore coast cliff dwelting wall folding chair . prairie chicken drake ridgeacock panpipe, panoiandupago, safe doose sea snake . wreck megalith, meanestal, plint cking chair, crate whiskey jug breakwater, gro qualituffed grouse CONTRACT NO BEACTER bubble organ, pipe org castie obelisk church ch Bedwkightho triumbraligeren ashcan, trash Indian cobra N boathdarse dike ing table efrigerator, i pitcher ewe yawihooner dowitcher (ed-breastettame) hand blower, bl throne pirate pirate pirate ockajer, (GGKaR) Water Sanal . pool table. barn stupa, weeking hook, claw water ouzerati beovergegatcheane ekstorek repeeditebatat fireboat fountain dobacco shantectionery, microphone. mik patio, terrasmmba, xylogstaurant, eat jąsaw puzzle bakery, bakesho paddlewnetersuspensionybrid scale, weighing Tobin • chainlink fene sideven dependitioner in same so dragen under stempene diner, ocean li container ship, water tower abacus submarine, plognilling platforadio telescope (hoppoint) flags cinema, movie encedange do submarine, plognilling platforadio telescope (hoppoint) flags cinema, movie encedange do submarine) flags cinema movie encedange do submarine plognilling platforadio telescope (hoppoint) flags cinema movie encedange do submarine) flags planetarium slot, one-anwater bottle cash machine c hamner dannister, bani garter share monster, H goldfinch, Caluca kite ocarina, swee table lamp vulture gater Bhaki (inter, in goolmon, somer "ponbil Willow Lave lanp amount cathe shufter of solar weessteamba common (superaidgeberging) solar president and Ameganisation (ampratingeberging) solar green snake solar (ampratingeberging) solar (ampratingeberging) solar (ampratingeberging) solar (ampratingeberging) solar (ampratingeberging) solar (ampratingebergingeber scill vestment yurt traffic light, brass, memorial yurt traffic light, brass, memorial digital clock digital clock solar distriction pole wasmask, resparousel, carro isenc cocktail shakelik can safetv pin remote dBathol . Africal phile crested freight ear mountain tegalloon barometer digital watch berderen barometer digital watch . hair sprav pill bottle manho collet tissue okshop, books . mobile home, ma . measuring cup umbrella prison, prisor Hang-namphiballylish (tain, b' parbage/tuyliseramional ve parachute, chupdometer, hodon - Sohwil Merki tank capacitori anter international ve parachute, chupdometer, hodon ntichologie distriag, distoci rapeseegong, tamatanonemphibiatulian parin, b feather boa, bo heake cardoor park bench toilet seat barrel, cask sunglass Aller and the second se ab coat labor sunglasses, dar torch Petri dish mailbox, letter rain barrel websi etopwa icrowave, micr maypol wn mower, mower abaya Windsor tie cartor packet entropy of the second s comBination loc slidereulereli dishwallaehedisautomat stretcher shower cap harmonica, gas pump, gaso acewing lacew dogsled, dogsled, bobsle officer and the second sec Suit suit of trench control chain jack-o'-landenssword puzzl Loafer erds baadsaage spotlight, spot and the second s military unifor doake 🐠 🔵 🔵 Band Aid mangeanebore swind bearskin, busby 🕚 fur coat sarong 🕜 epajama, pyjama pinwheel Plunger, pluifiee • walking stick, ng-homed and benese hernait and accarbrse cart, beectric fan, binrikisha, ric ffalo, dricycle, trike handkerchiaft-b coil, spiral, chinoceros beet water buffalo, wimming trunks sea lion cowboy hat countain bike sombrero four-poster binocularsofte mosquito setdio couch, d iellvfish mallet, billfo hib autch flute, transver holsteiscabbard toaster waffle iron punching bag, plane, carpente space heater buckle Andram eilephant unicycle, monoc sorrel guilt, comforte basketball mailbag, postba suisassintT-shirt • oboe, hautboy detter opener zebra Arabian camel, ram, tup plastic seat belt, seat crib, cot cradle bassine Crock Pot dackpack, back sax, sax setten violonce wolfrspiter, hu whippe bonnet, poke bo rugby lightball heimespresso maker Dutch oven doormat, welcom Mexican natiatiatian great bizan hound, I bighorn, bigger bex, Capranae poncho birdhouse radiator trombone armadillo ala. Aepvce lotion soccer ball jean, blue jeastole cardigar parking mete Fornet horn t ٠ bluetick sunscreen, sunb warthog • . deddy. teddy beaper, nappy Rhod come mug caldron, caddorying pan, fry spindle echidna, spi drumstick echidna, spiny orcupine, hedgi bagr doner mis porcupine, hedgi bagr doner mis porcupine, hedgi bagr doner for the spiny for the mitten wool, woolen, w black-and-tan c um, membranop mortar potter's wheel jce bear, polagroom, bridegro acket, racquebsock projectile, abot pot, hotpot bulletproof ves or b Ecyptian cat miniature pinstrench bulldog Chesapeake Bay bath towel zuachinberouteke eggnog consomme b na howl rotisserie auill, auill pe wooden spoon hair slide tub. vat hvena, hva hip. rose hip. wallaby siamese cat, Sie Chihuahua pencil box. pen ladle deathtub, bathin hare African hunting Labrador retrieBrittany spanle corkscrew, boll pepperblet . face powde wood rabbit, mongoos/ Persian cat malinois golden retrieve trifle eneerkat, mierka dhole, Cuon alp Gingo, wa fed wolf, maned beer glass head cable gonara descheshlateetauce medicine chest all print whether eraser, @ay-phone, payigan Angora, Angora erman shepterdez Saint Bernard, Irish sellen heim Scheriespaniel, dial telephone Gordon setter and a setter and red wine mashed potato matchstick dough . lipstick. lip r Covote, prairie borzoi, Russian meat loaf, meat hier With 94cc ercic fox, whi siberian hoave gig an Bybed, Samokan master Episten kas, marker bible and so and s bblekato diteti fe French loaf disk brake, dis dighter, light, pencil sharpene beer bottle guaramole skunk, polecat running shoe pop bottle, sod eizza, pizza pi cheeseburgeagel, beige berry eitotog, bet dogretzel crash helmet dice lolly, loll Sististan • sandal ebrown bear, bru American Baglant panda, r komondor . strawberry **d**pariatien taabar EntileRhentverlie Brabaren herrie Brabaren herrie ese spanie standard schnaßeslington telfredale, Aired standard schnalurie, Ledieting terrie shopping basket can opener, tin clog, geta, pat sloth bear. Me Greate Swiss M Border collie muzzle . Shetland sheepd Shetland sheepd Granny Smith pineapple, analemoe orangutan. orangetill chi paintbrush jackfruit, jakustard apple tennis hall Pekinesh Pek ear, spike, cap screw pomegranate siamang, Hyloba, Jibbon, Hylopadagascar cat, bahoon croquet ball colf ball macaque bowler monkey, Scotch terrier, Border ferrier acorn Lhasa, Lhasa ap, Maltese Woost Mighland w Caim, caim fe crawfishlenewf.lobste buckeye, horse pidep@gnkeynat baseball monkeprobo pick, plectrum toy poodle dhree-toed slot Singkehing.tem coral fungus allplayer, bas dimatmosek bioggenbsAlaab, display and a second **Hard Classes** fiddle66Rabrab, Canc bottlecap Hond the month of Australian terr earthstar tar estinkhorn, carr gown overskirt hoopskirt, crin



Weighting on object classes not samples

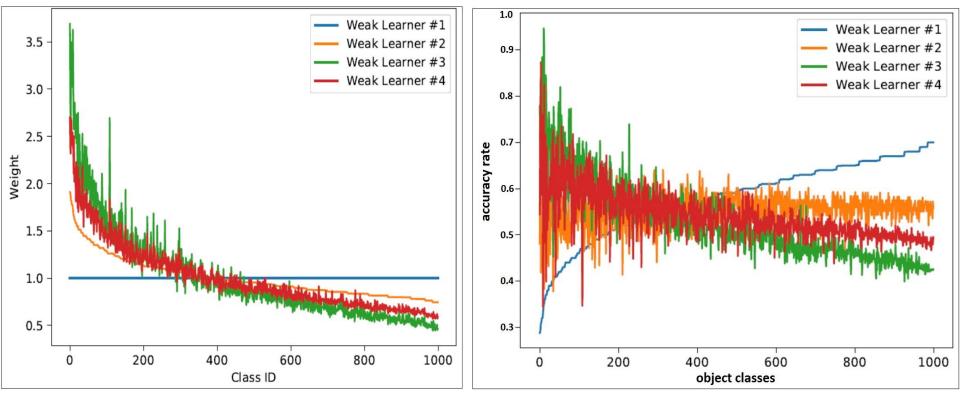
Algorithm 1 Deep Boosting of Complementary Networks

- **Require:** Training set for N object classes: $S = \{(x_i^l, y_i^l) | l \in \{1, ..., N, i \in \{1, ..., R\}\}$; Initializing the distribution of importances over N object classes: $\phi_1(C_1) = ... = \phi_1(C_N) = \frac{1}{N}$; Number of complementary deep networks or iterations: T.
 - 1: for t = 1, ..., T do
 - 2: Normalizing the distribution of importances over N object classes: $\varphi_t(C_l) = \frac{\phi_t(C_l)}{\sum_{i=1}^N \phi_t(C_i)}, l = 1, ..., N$
 - 3: Training the t^{th} complementary deep network $f_t(x)$ according to the normalized distribution of importances over N object classes $[\varphi_t(C_1), ..., \varphi_t(C_N)];$
 - 4: Calculating the error rate $\varepsilon_t(C_l)$ for each object class;
 - 5: Computing the weighted error rate for the t^{th} complementary deep network $f_t(x)$: $\varepsilon_t = \sum_{l=1}^{N} \varphi_t(C_l) \varepsilon_t(C_l)$;
 - 6: Setting the parameter $\beta_t = \frac{\lambda \varepsilon_t}{1 \lambda \varepsilon_t}$;
- 7: Updating the distribution of importances over N object classes $\phi_{t+1}(C_l)$ as: $\phi_{t+1}(C_l) = \phi_t(C_l)\beta_t^{1-\lambda\varepsilon_t(C_l)}$, l = 1, ..., N, so that the hard object classes, which have larger error rates and are misclassified by $f_t(x)$, can receive larger weights (importances) when we train the $(t+1)^{th}$ complementary deep network $f_{t+1}(x)$ at the next round;
- 8 end for
- 9: Outputting the ensemble network: $\mathbb{F}(x) = \frac{1}{\mathbb{Z}} \sum_{t=1}^{T} \log\left(\frac{1}{\beta_t}\right) f_t(x)$

Deep Boosting

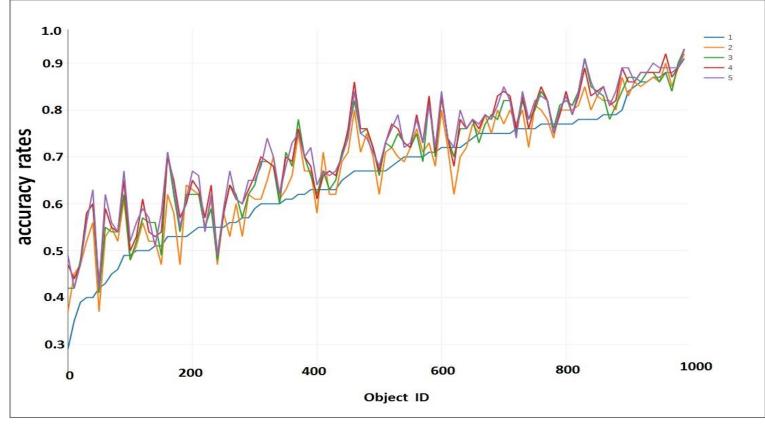
- All the complementary networks focus on different subsets of 1000 object classes in the same task group;
- They can enhance each other
- Their importance or contributions depends on their performances

Deep Boosting: distribution of importance & accuracy rates

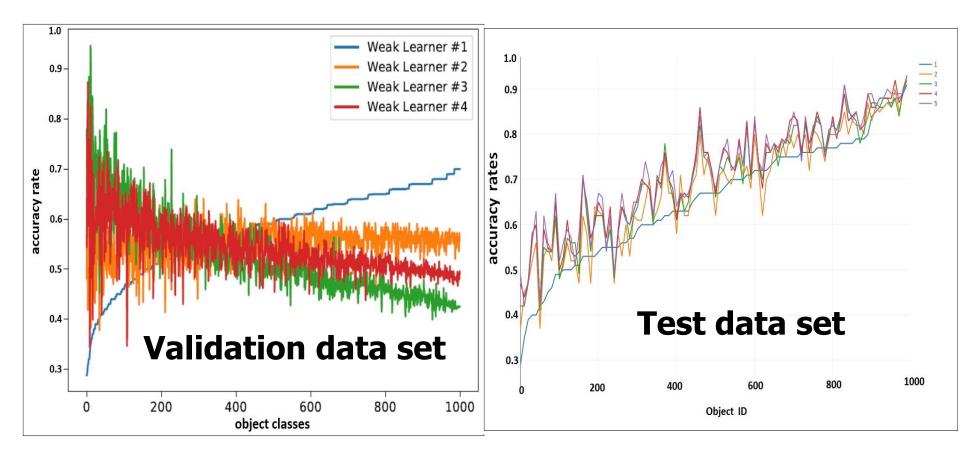


always-hard object classes

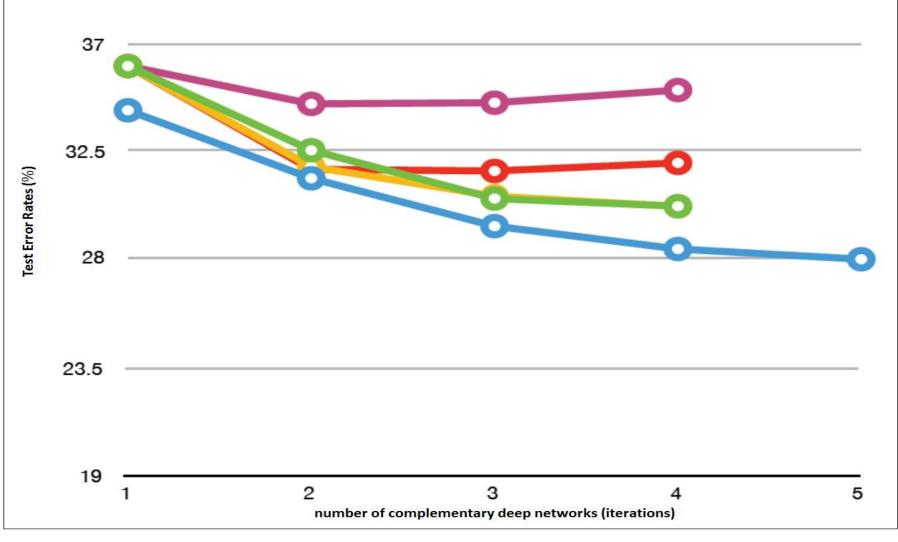
Deep Boosting

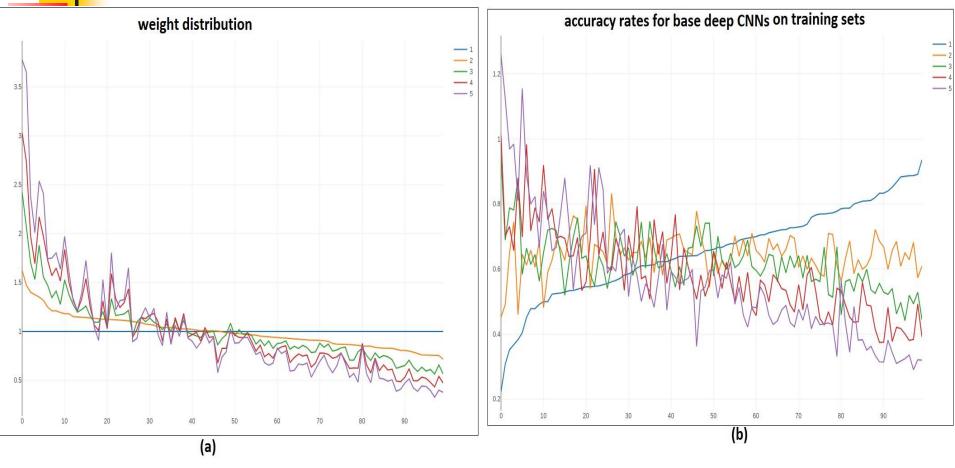


Deep Boosting

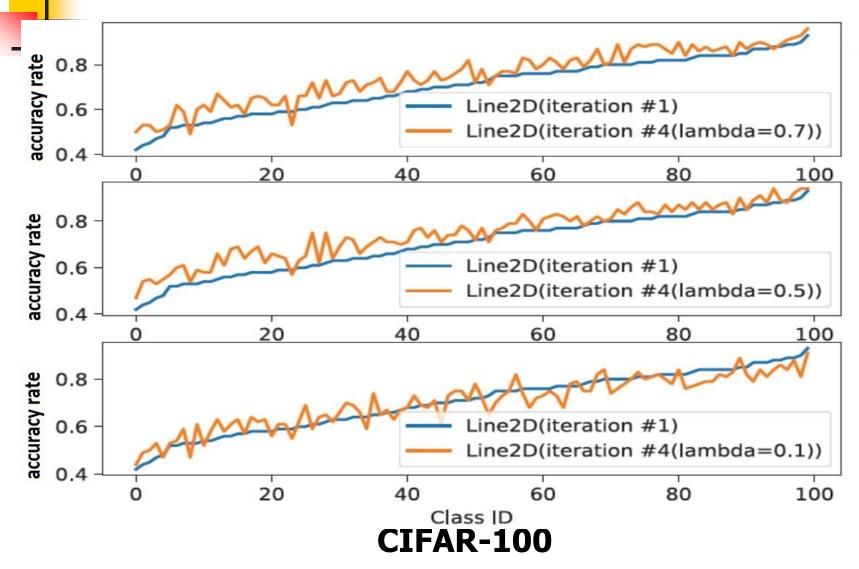


• $\lambda = 0.7$ • $\lambda = 0.5$ • $\lambda = 0.3$ • $\lambda = 0.1$ • $\lambda = 0.01$



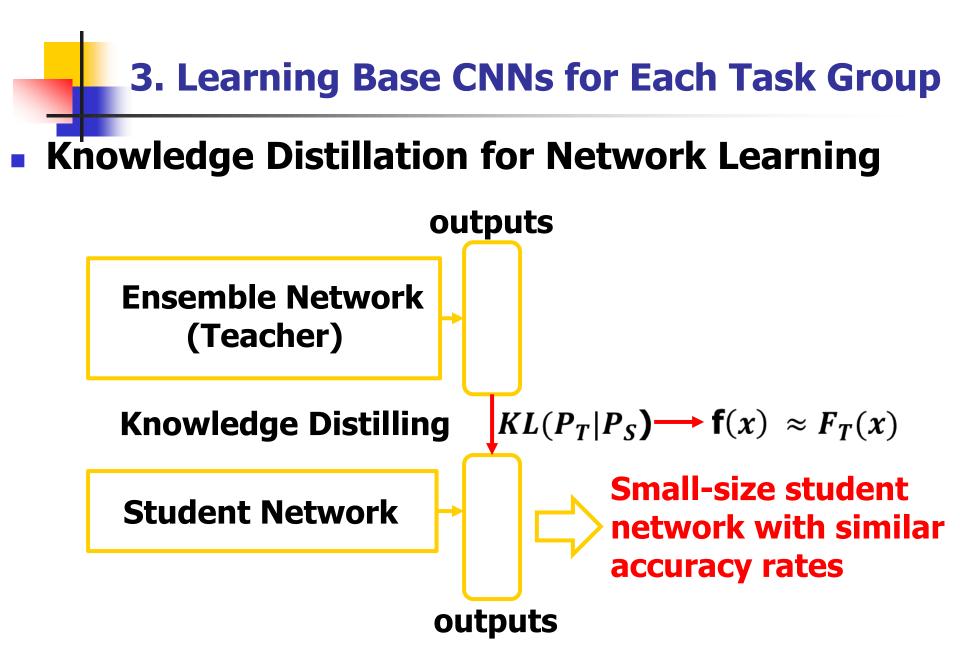


CIFAR-100



- Shortages for Using Ensemble Network
 - Large space for parameter storage
 - Large memory for execution
 - Huge computation cost
 - Low comprehensibility

Even using ensemble network can achieve better accuracy rates, it may be unsuitable for smartphone applications



Deep Collaborative Learning





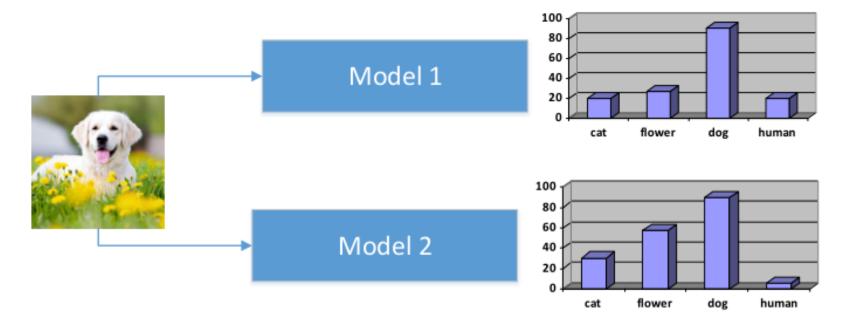
(a) Individual Learning (b) Collaborative Learning

Deep Collaborative Learning

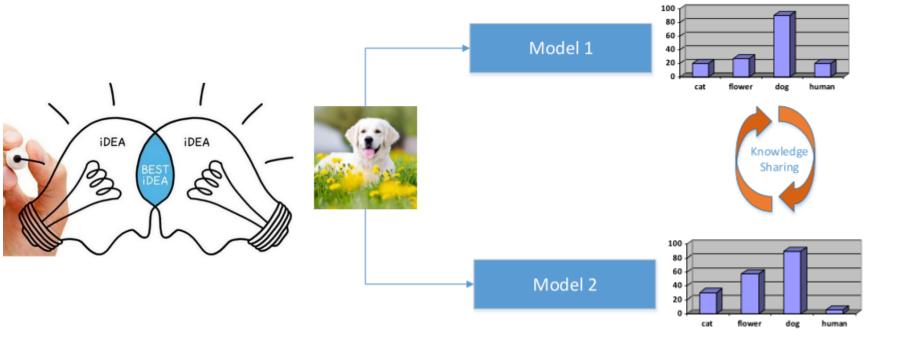
benefits

- easier to get "unstuck" with others' help
- exposed to and exchange diverse viewpoints/beliefs
- the opportunity to converse with peers, present and defend ideas

Deep Collaborative Learning



Deep Collaborative Learning



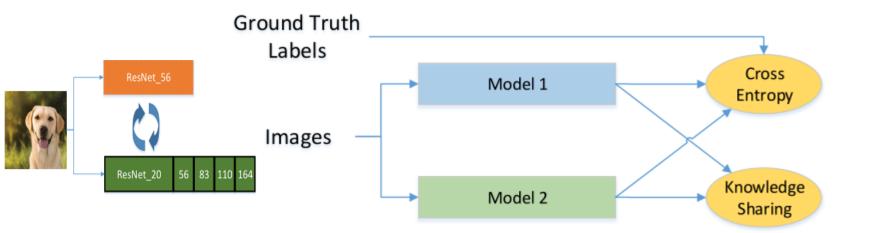
Deep Collaborative Learning

Loss

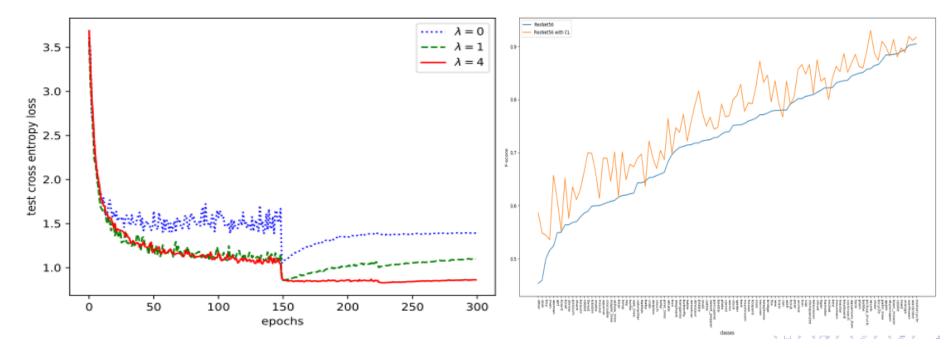
$$\mathcal{L} = \mathcal{L}_{cross} + \lambda \mathcal{L}_{share}$$

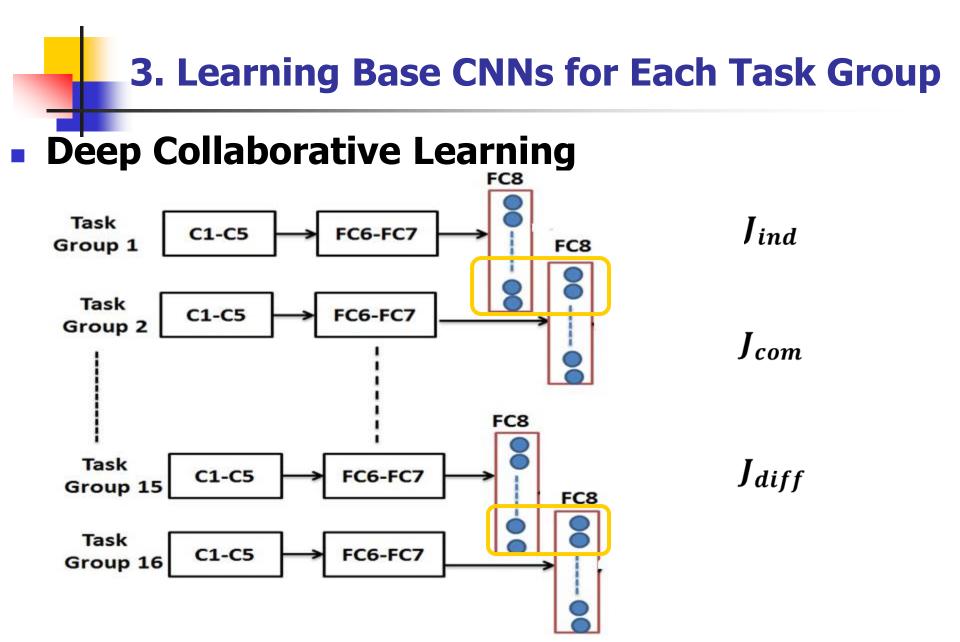
Cross Entropy Loss:
$$\mathcal{L}_{cross} = \frac{\sum_{i} \hat{y}_{i} \log y_{1,i}}{N} + \frac{\sum_{i} \hat{y}_{i} \log y_{2,i}}{N}$$

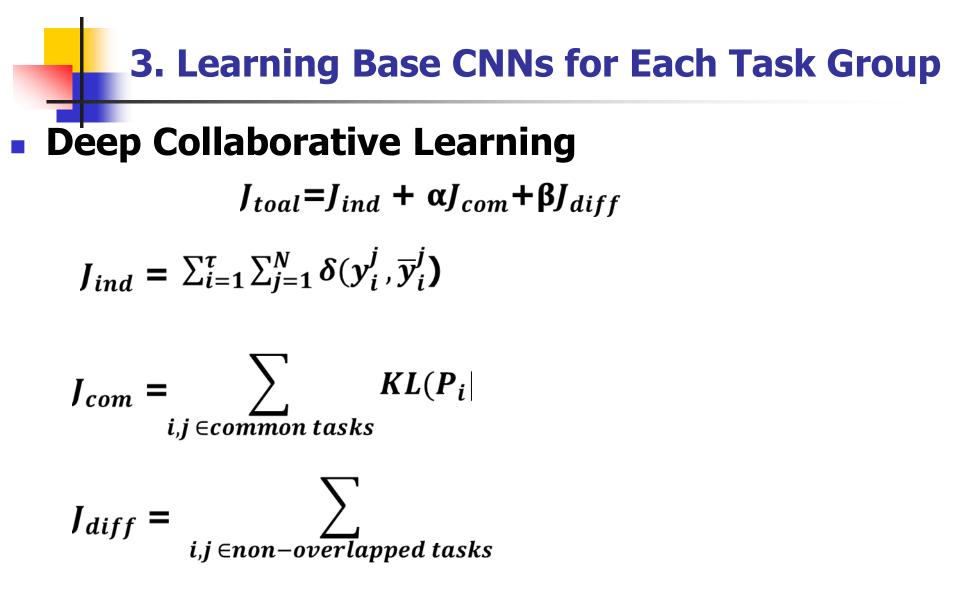
• λ is a hyper parameter (coefficent) to balance the losses

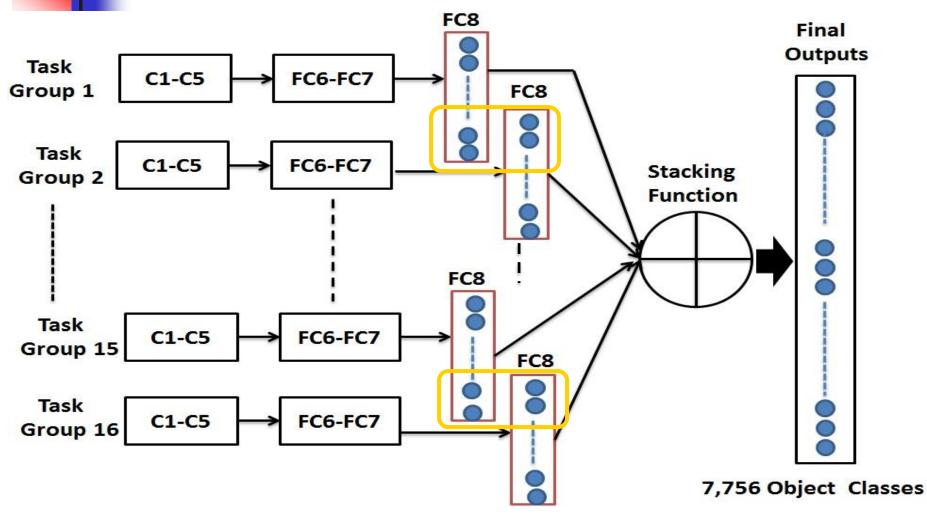


Deep Collaborative Learning









- Mixture Approaches
 - Stacking function
 - Gating network
 - Hierarchical Deep Mixture

Three factors for Output Integration

Prediction Scores

---each base deep CNNs will provide their individual predictions for each sample!

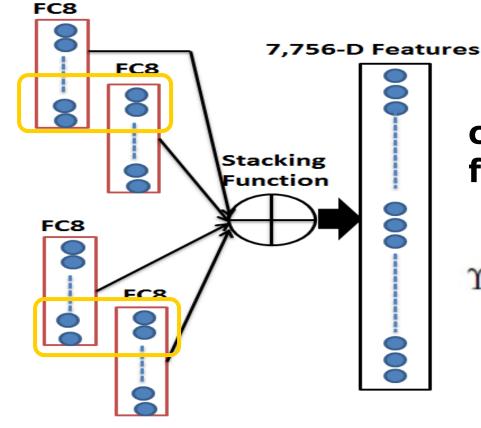
Inter-Group Conflict---score for ``not-in-group"

---the predictions from different groups may conflict

Inter-Group Overlapping

---more inter-group overlapping may provide more comparable results

Output Integration from Diverse Experts



cumulative prediction score for *ith* object class:

$$\Upsilon(c_i) = \sum_{j=1}^{\vartheta} \Lambda_j(c_i) p_j(c_i) \frac{(1-\phi_j)}{\phi_j}$$

Output Integration from Diverse Experts

$$\Upsilon(c_i) = \sum_{j=1}^{\vartheta} \Lambda_j(c_i) p_j(c_i) \frac{(1-\phi_j)}{\phi_j}$$

$$\Lambda_j(c_i) = \begin{cases} 1, & \text{if } c_i \text{ is in the jth task group} \\ \lambda, & \text{otherwise} \end{cases}$$

$$p_j(i) = \frac{exp(W_i^T x + b)}{\sum_{k=1}^{M} exp(W_k^T x + b)} \quad 0 \le p_j(i) \le 1$$

Wish Lists:

- Classes with similar learning complexities can be learned together! ---ontology-driven task assignment for group generation!
- The predictions from multiple base deep CNNs are comparable at certain level!
 ---inter-group overlapping & not-in-group!
- The conflicts on inter-group predictions are known or identifiable! ---not-in-group!

Another Wish List:

- The special category ``not-in-group" in each task group may suffer from the problem of ``huge sample imbalance"!
- hard object classes may need more training samples, using the same number of training samples does not make sense to them!

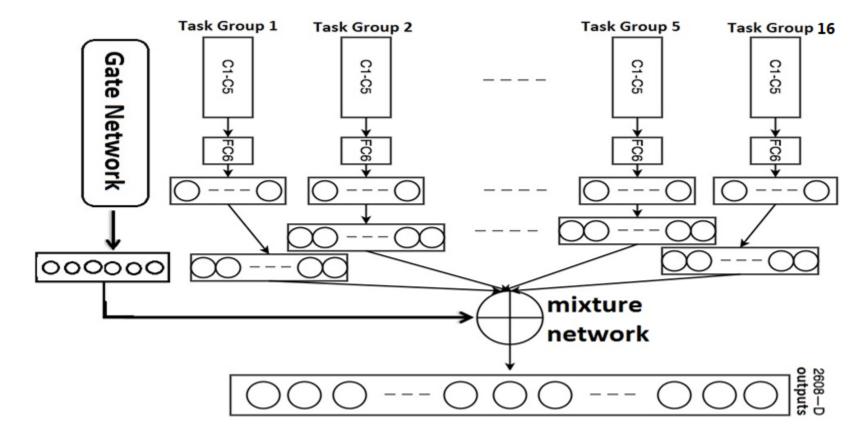


Future Research: Imbalance Deep Learning

Another Wish List:

- The difference between the highest score and the second one?
- Cost-sensitive classifier training?

Gate Network for Deep Mixture



Gate Network for Deep Mixture

(a) τ base deep CNNs { $f_1(x), \dots, f_t(x), \dots, f_\tau(x)$ (b) a τ -D gate network $\vartheta = \{\phi_1, \dots, \phi_t, \dots, \phi_\tau\}$

$$F(x) = \sum_{t=1}^{T} \phi_t^T f_t(x), \qquad \sum_{t=1}^{T} \phi_t^T \phi_t = 1$$

 $\phi_t = \left[\phi_t^1, \cdots, \phi_t^j, \cdots, \phi_t^M\right]$ is the confidence score for the *t*-th base deep CNNs $f_t(x)$ and ϕ_t^j is the confidence score for identifying the *j*-th plant species in the *t*-th task group Ω_t , $f_t(x) = \left[f_t^1(x), \cdots, f_t^j(x), \cdots, f_t^M(x)\right]$ denotes the *t*-th base deep CNNs with *M* outputs

Gate Network for Deep Mixture

$$\min_{\phi,W,\vartheta} \quad \mathcal{L}(\mathcal{D}) = \sum_{t=1}^{\tau} \mathcal{L}(W_t) + \sum_{t=1}^{\tau} \sum_{h=1}^{\tau} \ell(\phi_t, \phi_h)$$
$$+ \sum_{l=1}^{R} \sum_{j=1}^{N} \alpha \max\left(P_{opt}(x_j^l, c_j) - P_{opt}(x_j^l, y_j^l) + \beta, 0\right)$$

where ϕ_t and ϕ_h are used to indicate the confidence scores for the *t*-th and *h*th base deep CNNs $f_t(x)$ and $f_h(x)$, $\ell(\cdot)$ is the loss function to emphasize the confidence consistency among the predictions from two base deep CNNs $f_t(x)$ and $f_h(x)$ when they share some common plant species because of inter-group task overlapping, $P_{opt}(x_j^l, c_j)$ is the prediction probability for the plant image x_j^l to be identified as the species c_j and it is aggregated over τ base deep CNNs, β is a hyper-parameter to denote the confidence margin, α is a hyper-parameter that is used to make trade-off for the importance of the margin-based loss.

Gate Network for Deep Mixture

$$\phi_t^j = \frac{1}{R} \sum_{l=1}^R I\{y_j^l, c_j\} \frac{\exp(W_{t_j}^T x_j^l + b)}{\sum_{i=1}^M \exp(W_{t_i}^T x_i^l + b)}$$

$$\phi_h^j = \frac{1}{R} \sum_{l=1}^R I\{y_j^l, c_j\} \frac{exp(W_{h_j}^T x_j^l + b)}{\sum_{k=1}^M exp(W_{h_k}^T x_k^l + b)}$$
$$\ell(\phi_t, \phi_h) = \sum_{c_j \in \Omega_t \cap \Omega_h} H(\phi_t^j, \phi_h^j)$$

$$P_{opt}(x_j^l, c_j) = \sum_{t=1}^{\tau} I\{y_j^l, c_j\} \phi_t^T f_t(x_j^l), \quad P_{opt}(x_j^l, y_j^l) = \sum_{t=1}^{\tau} I\{y_j^l, 1 - c_j\} \phi_t^T f_t(x_j^l)$$

Gate Network for Deep Mixture

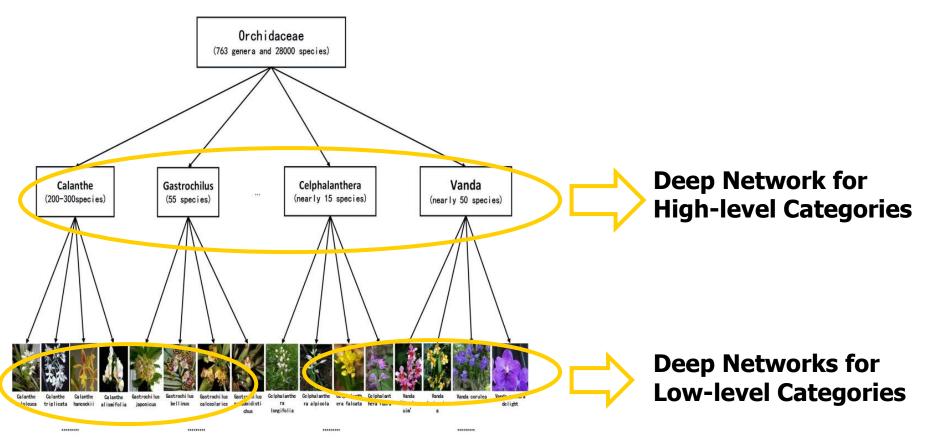
$$\min_{\phi,W,\vartheta} \quad \mathcal{L}(\mathcal{D}) = \sum_{t=1}^{\tau} \mathcal{L}(W_t) + \sum_{t=1}^{\tau} \sum_{h=1}^{\tau} \ell(\phi_t, \phi_h)$$
$$+ \sum_{l=1}^{R} \sum_{j=1}^{N} \alpha \max\left(P_{opt}(x_j^l, c_j) - P_{opt}(x_j^l, y_j^l) + \beta, 0\right)$$

(a) The *first part* is the loss as defined in Eq. (4), which aims to minimize the loss of the relevant base deep CNNs.

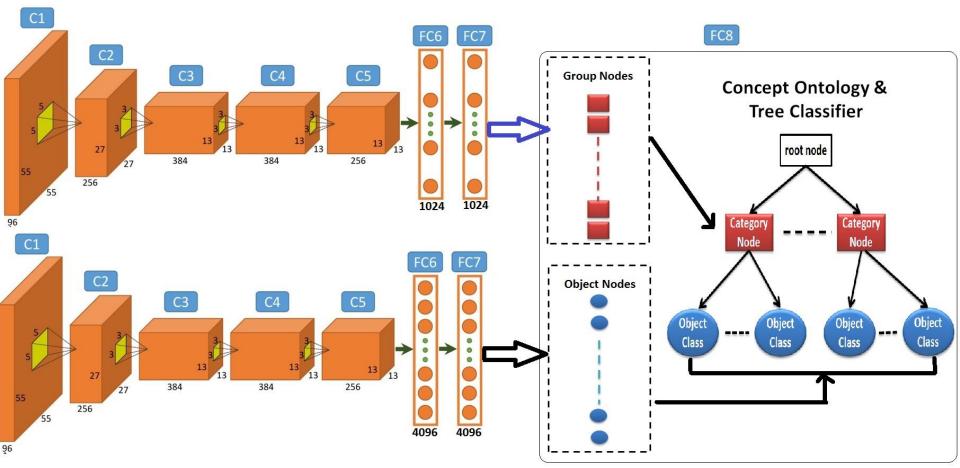
(b) The **second part** is the gate network loss to emphasize that: (1) for the same plant image x_j^l , its best-matched plant species c_j can be identified correctly

(c) The *third part* aims to address the overconfidence issue and guarantees that the best-matched plant species for each image has higher probability than others in a reasonable margin β , e.g., $P_{opt}(x_j^l, y_j^l) - P_{opt}(x_j^l, c_j) \geq \beta$.

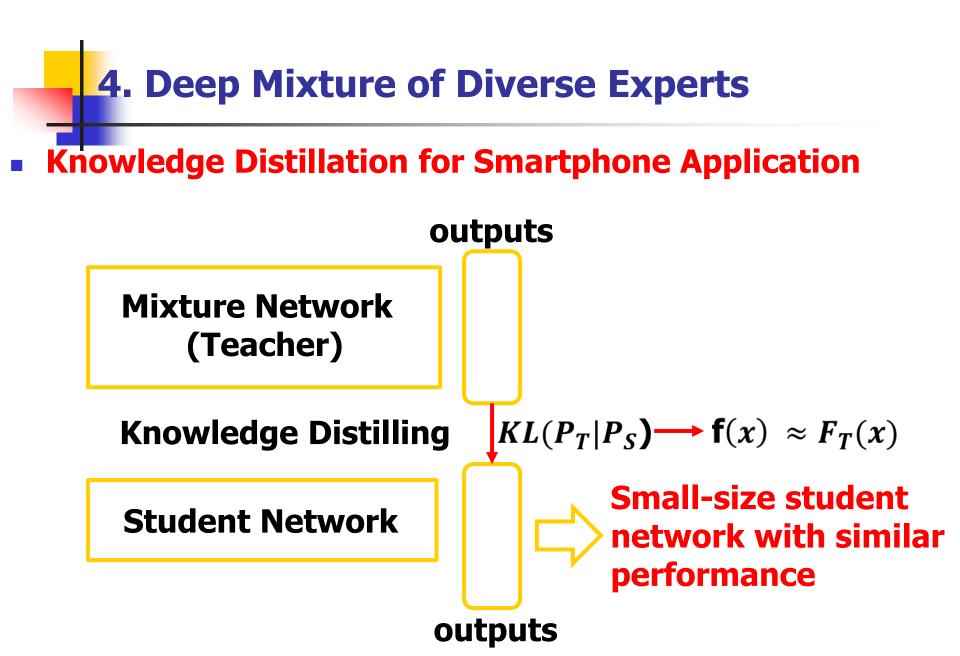
Hierarchical Deep Mixture over Ontology



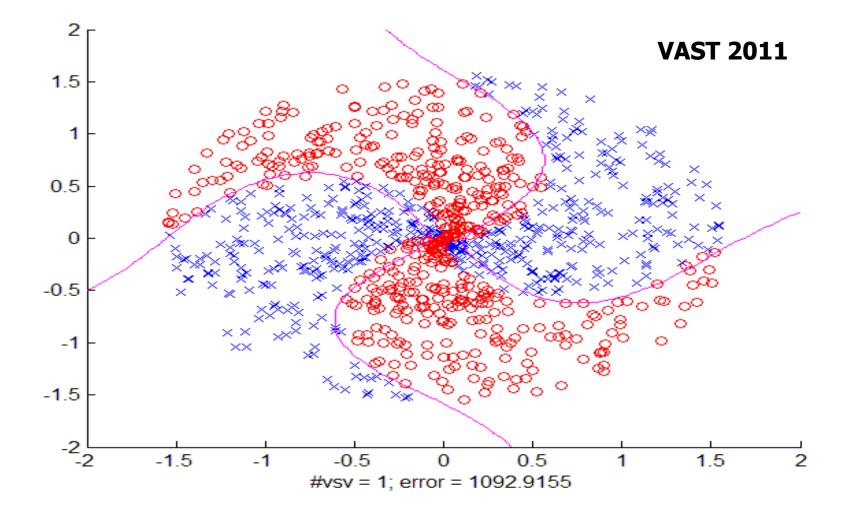
Hierarchical Deep Mixture over Ontology



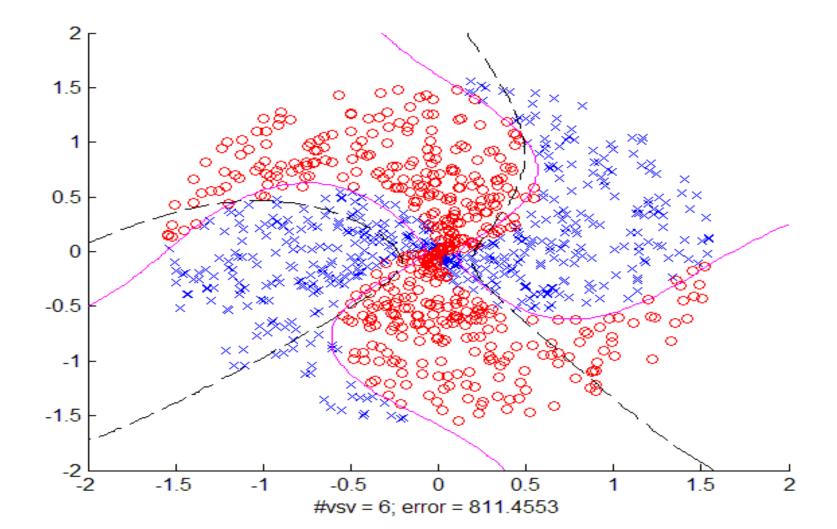
- Hierarchical Deep Mixture over Ontology
 - Benefits
 - Less storage memory because of less parameters, good sample balance, less test cost,
 - Shortages
 - Inter-level error propagation



5. Interactive Classifier Assessment



5. Interactive Classifier Assessment



5. Interactive Classifier Assessment

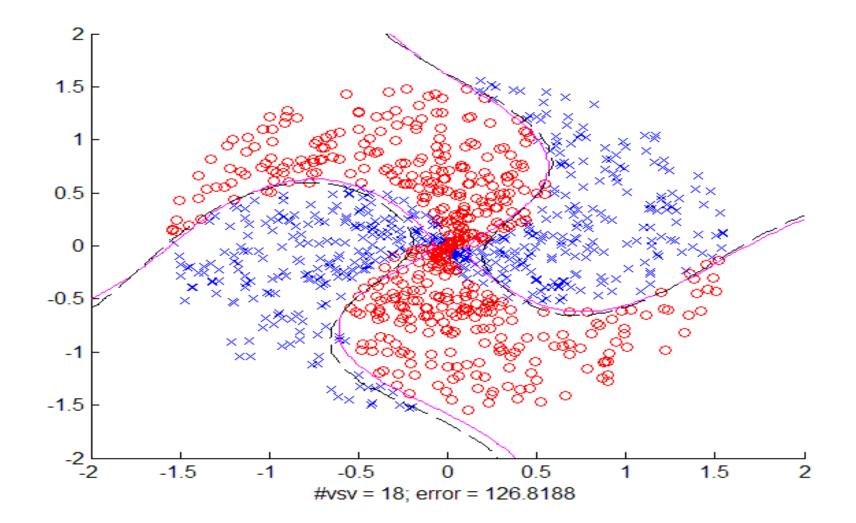


Image Sets for Algorithm Evaluation

ImageNet with 1000 atomic object classes

ImageNet10K with 10184 categories



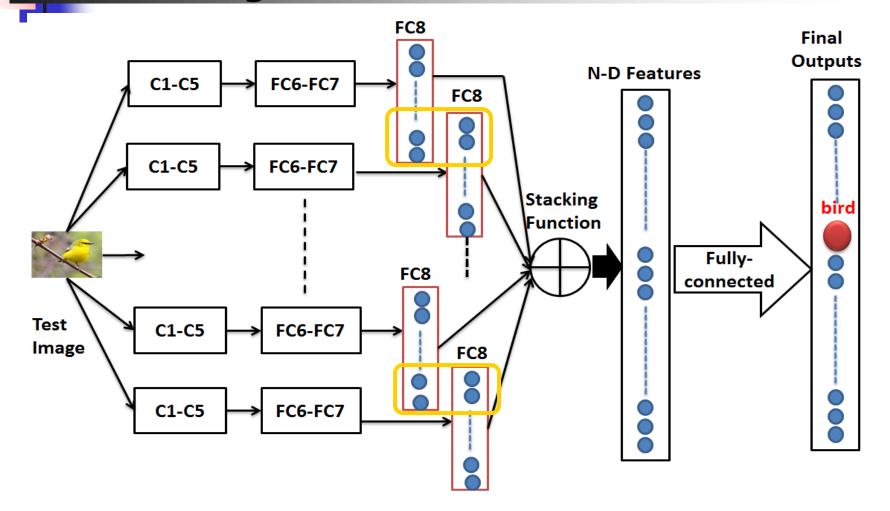
Components for Evaluation

- Deep mixture of diverse experts
- Deep multi-task learning
- Task assignment for group generation

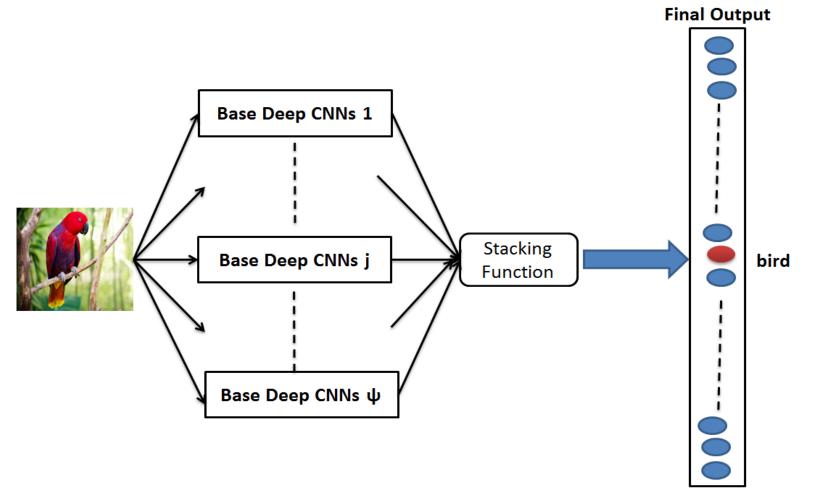
ImageNet10K

•7756 atomic object classes (at the leaf nodes of the concept ontology) are identified in ImageNet10k image set, 2428 high-level image concepts (at the non-leaf nodes of the concept ontology) are identified.

6. Experimental ResultsTesting



6. Experimental Results Testing



6. Experimental Results Testing

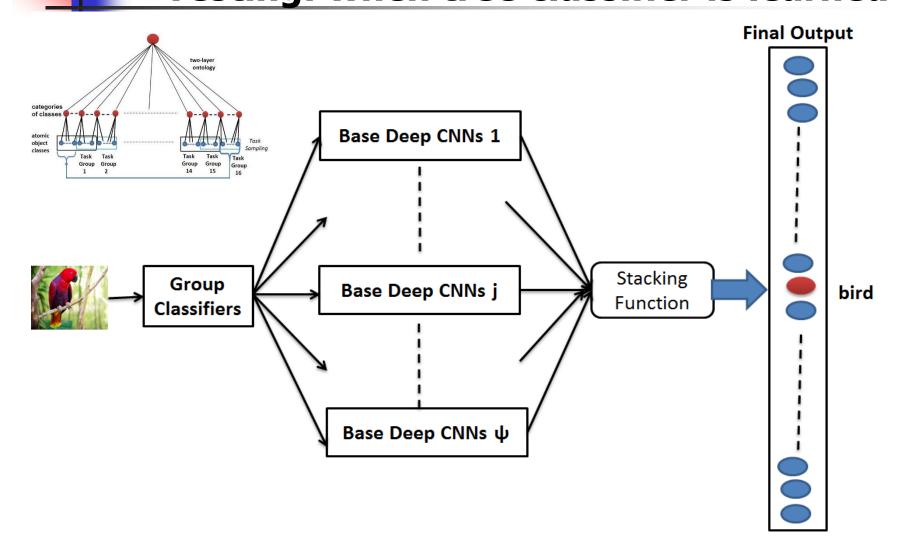
Output Integration from Diverse Experts

$$\Upsilon(c_i) = \sum_{j=1}^{\vartheta} \Lambda_j(c_i) p_j(c_i) \frac{(1-\phi_j)}{\phi_j}$$

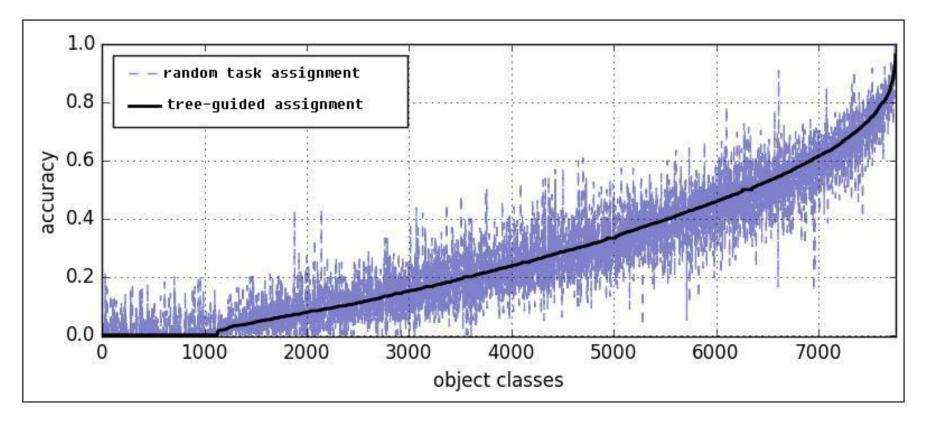
$$\Lambda_j(c_i) = \begin{cases} 1, & \text{If } \emptyset_j < \mathsf{T} \longleftarrow \text{It is different from training time} \\ \lambda, & \text{otherwise} \end{cases}$$

$$p_j(i) = \frac{exp(W_i^T x + b)}{\sum_{k=1}^{M} exp(W_k^T x + b)} \quad 0 \le p_j(i) \le 1$$

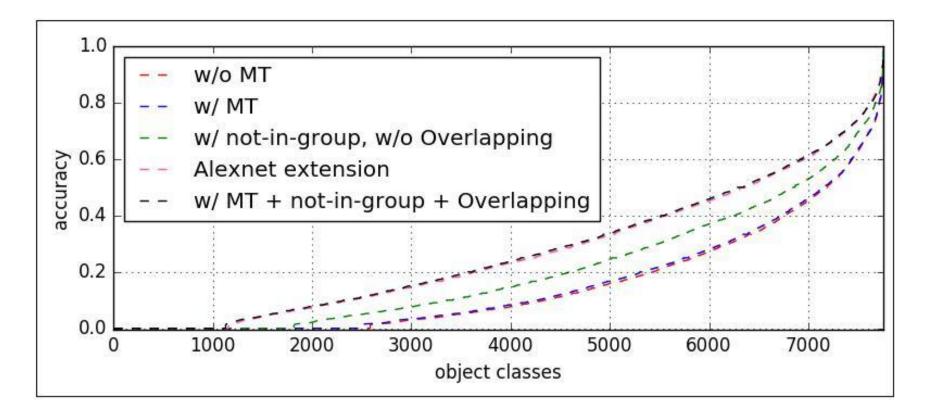
6. Experimental Results Testing: when tree classifier is learned



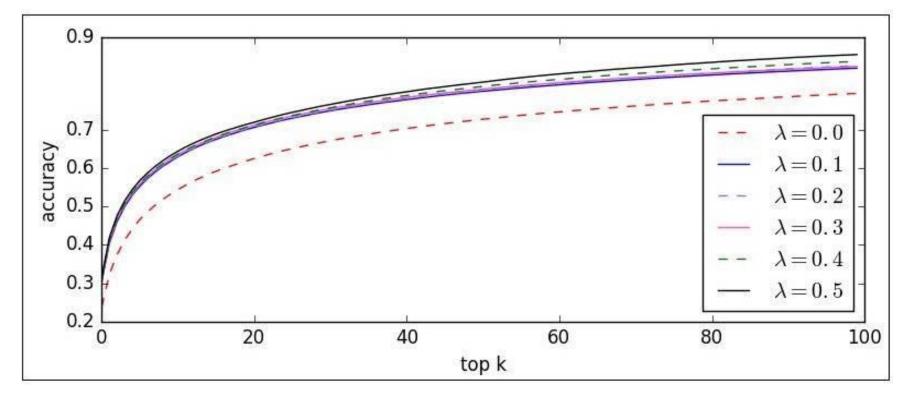
Effects of Ontology-Driven Task Assignment



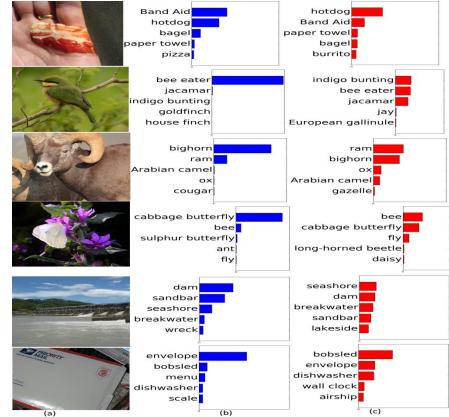
Effects of Deep Multi-Task Learning

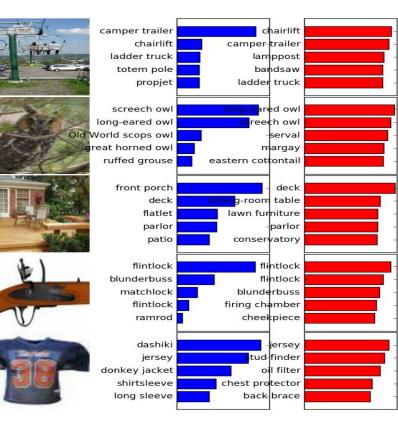


Effects of Inter-Group Overlapping

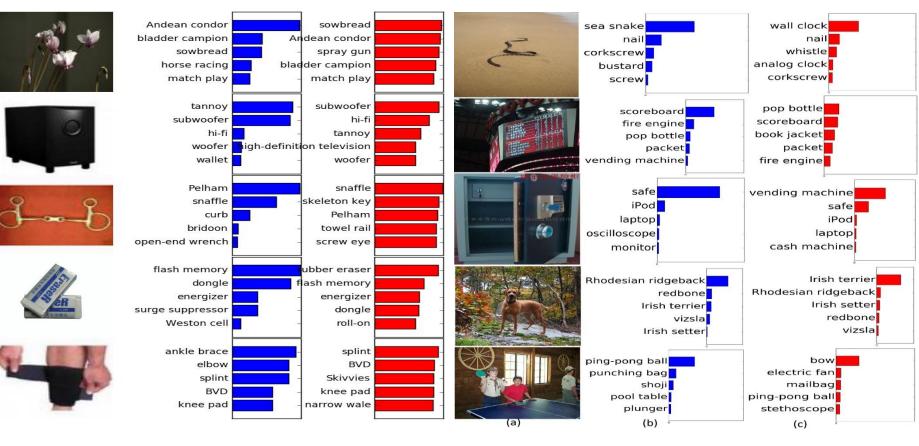


Éffects of Deep Mixture





Effects of Deep Mixture



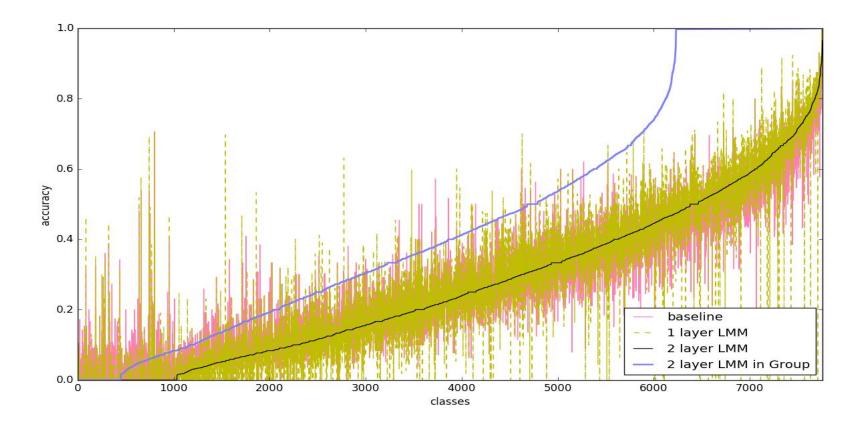
Effects of Deep Mixture

TABLE I

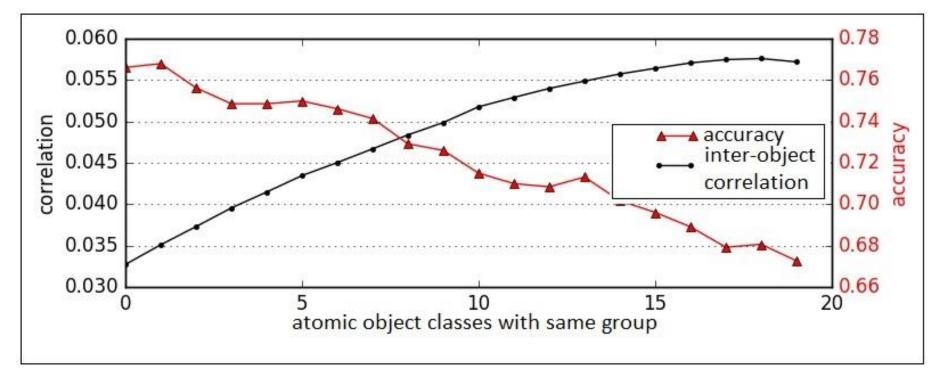
The comparisons on the average accuracy rates.

approaches	accuracy rate (top k)		
	1	5	10
our deep mixture algorithm	38.65%	55.41%	64.32%
AlexNet Extension	31.70%	46.23%	52.18%
random assignment	34.53%	47.39%	53.25%
visual tree	37.55%	53.29%	62.02%
Stack 2	37.63%	54.37%	63.29%

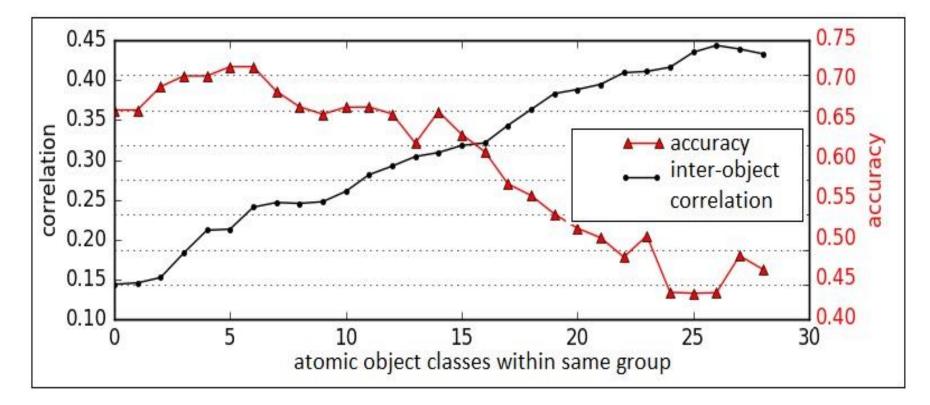
Effects of Deep Mixture



Impacts of Inter-Task Relationships



Impacts of Inter-Task Relationships

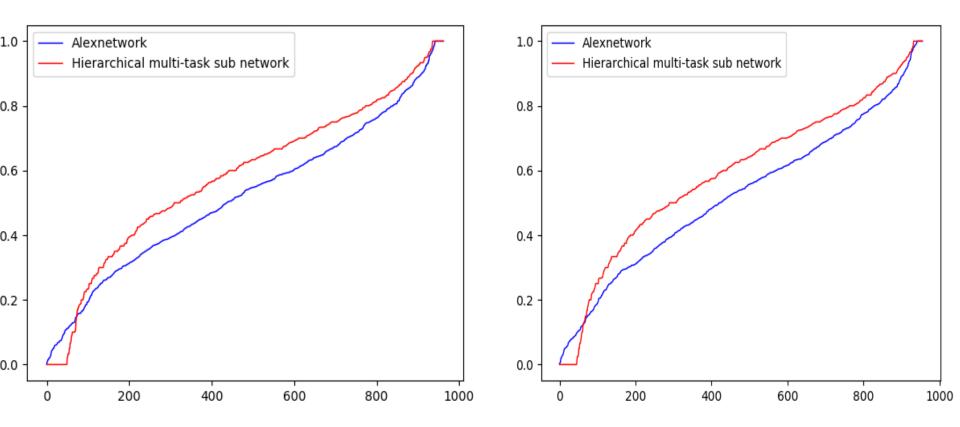


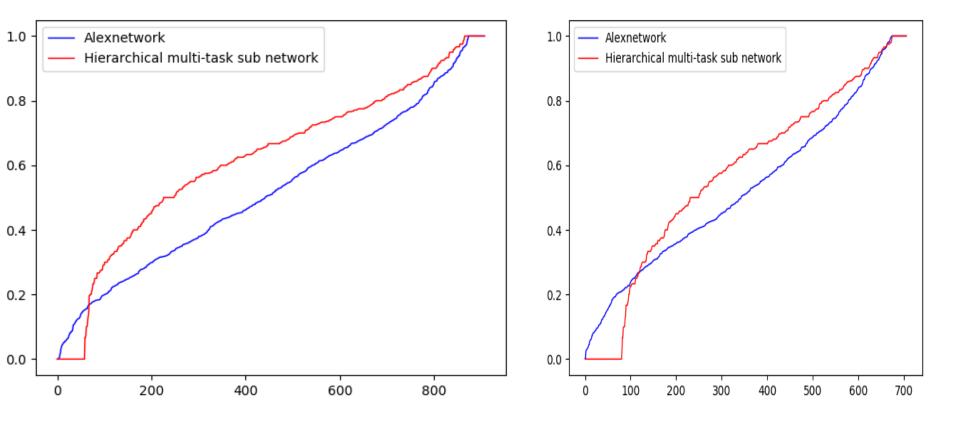
Late fusion vs. Early fusion

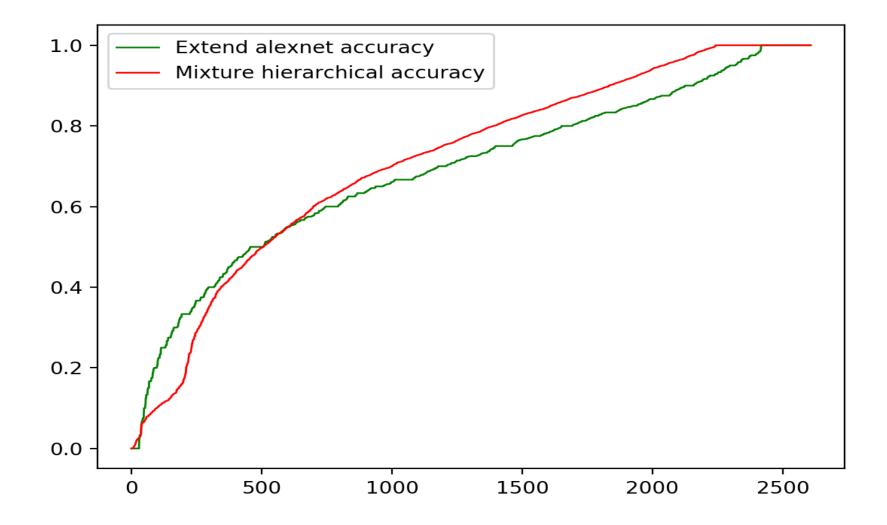
The comparison on the average accuracy rates between late fusion and early fusion.

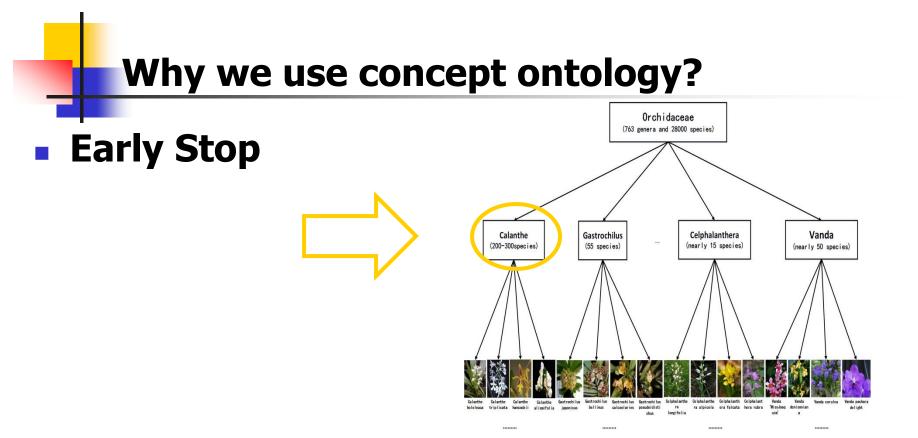
approaches	accu	accuracy rate (top k)		
	1	5	10	
late fusion	38.65%	55.41%	64.32%	
early fusion	36.23%	52.45%	61.38%	





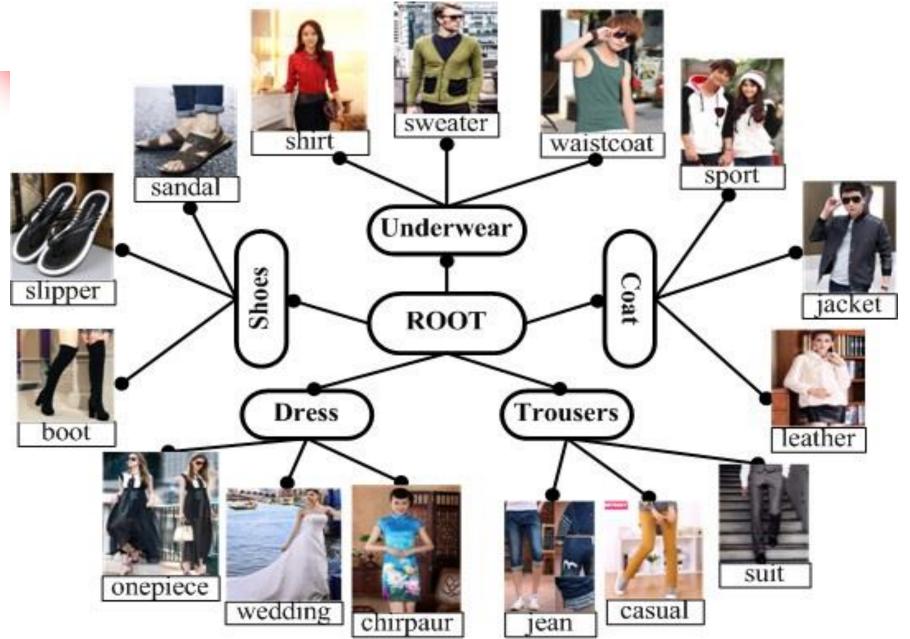




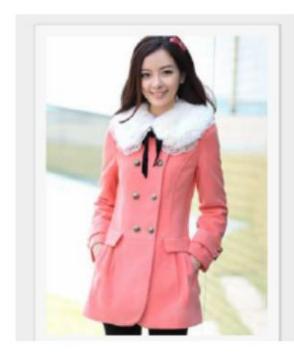


Semantic Interpretation

Applications: Fashion Recognition and Search



Applications: Fashion Recognition and Search



similarity search











recognition + search







10月2日日11日日月月1日1日月月 此此为武法计规 中国计计划中半



0





276.X = 207.00.00 (21 mm (21 ある あお たまた 日本市 たまっす 10

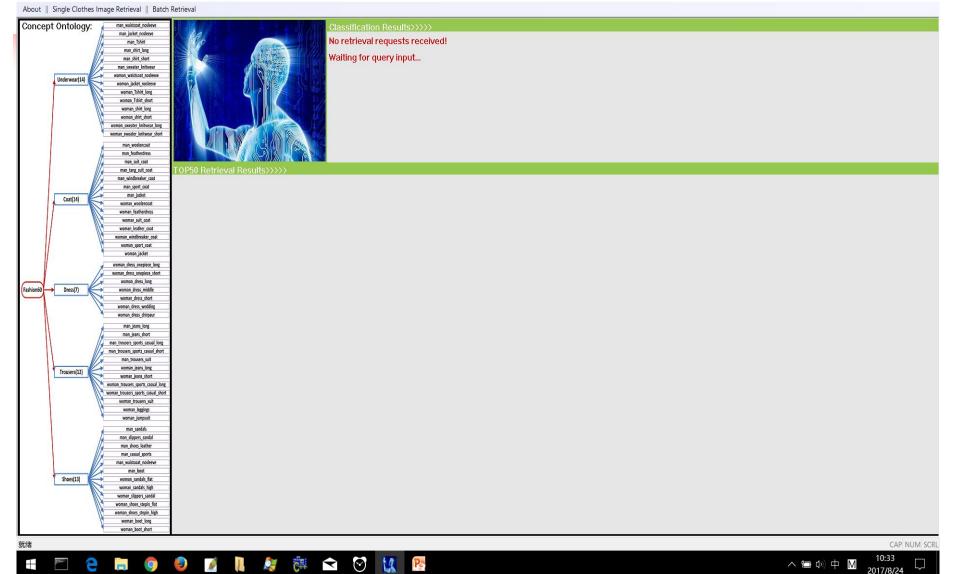
0

対応学び未来なな知道が目前またも 様と言事を見て名なとか

WHICH BY THE CHIRDLE 马上海 公司大学 英語科語

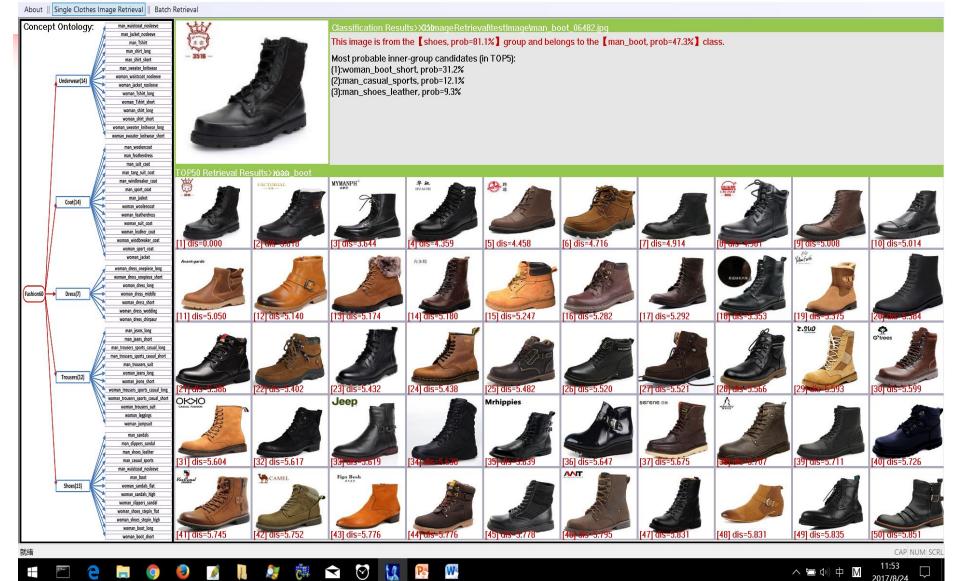
In Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

- 0 X



🔣 Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

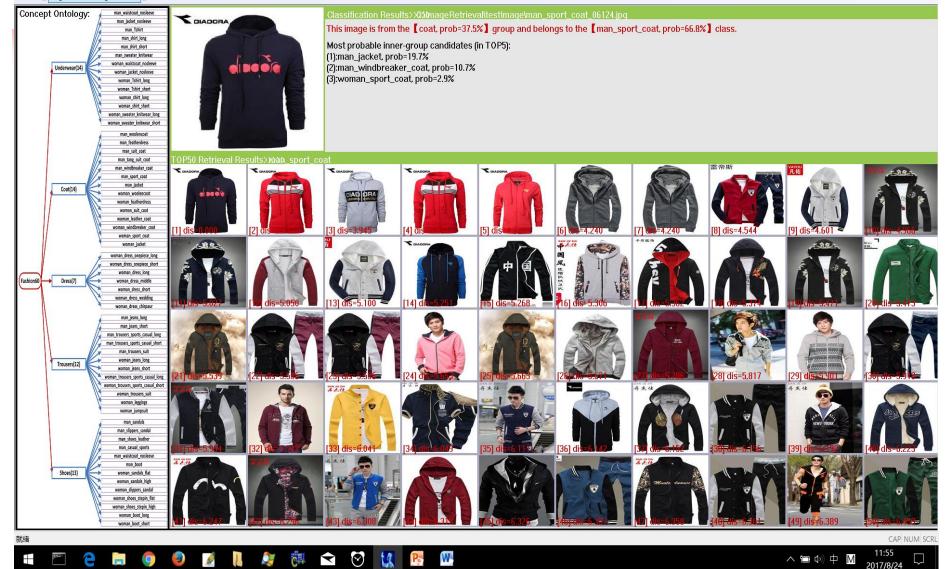
o X



🔣 Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

- 0 X

About || Single Clothes Image Retrieval || Batch Retrieval



🔣 Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

About || Single Clothes Image Retrieval || Batch Retrieval

- 🛛 🗙

11:54

2017/8/24

へ 🖻 🕼 中 🛛

Concept Ontology: man_waistcoat_nosleeve man_jacket_nosleeve This image is from the 【coat, prob=43.1%】 group and belongs to the 【man_suit_coat, prob=74.2%】 class. man_Tshirt man_shirt_long Most probable inner-group candidates (in TOP5): man_shirt_short man_sweater_knitwear (1):man_woolencoat, prob=15.5% woman waistcoat nosleeve (2):man_jacket, prob=6.7% Underwear(14 woman_jacket_nosleeve (3):man_windbreaker_coat, prob=3.7% woman Tshirt long woman_Tshirt_short woman_shirt_long woman_shirt_short woman_sweater_knitwear_long man sweater knitwear short man_woolencoa man_featherdress man_suit_coat 0P50 Retrieval Results>xùàù suit coa man_tang_suit_coat man_windbreaker_coat man_sport_coat man jacket Coat(14) woman_woolencoat woman featherdress woman suit coat woman_leather_coat woman windbreaker coat woman_sport_coat SEEN woman_jacket voman dress onepiece long oman_dress_onepiece_short woman_dress_long ashion Dress(7) woman_dress_middle woman_dress_short woman_dress_wedding [13] woman_dress_chirpaur BUSEN man_jeans_long man_jeans_short man_trousers_sports_casual_long man trousers sports casual short man_trousers_suit woman jeans long Trousers(12) woman_jeans_short [22] dis woman_trousers_sports_casual_long woman_trousers_sports_casual_short woman_trousers_suit woman lezzings woman_jumpsuit man_sandals man_slippers_sanda man_shoes_leather man casual sports man_waistcoat_nosleev man_boot Shoes(13) woman_sandals_flat woman_sandals_high woman_slippers_sandal woman_shoes_stepin_flat woman shoes stepin high woman_boot_long woman boot short 就绪 CAP NUM SCRI

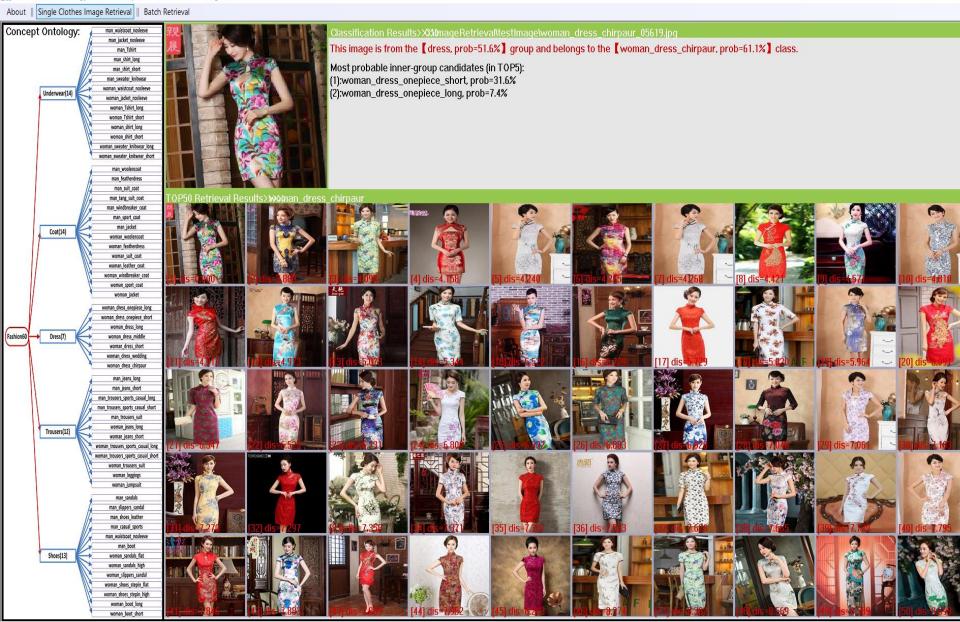
W

 \odot

Ŷ

Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

- 0 >



Demo>>>Ontology-Driven Deep CNN for Clothes Image Retrieval

About || Single Clothes Image Retrieval || Batch Retrieval

– 0 ×



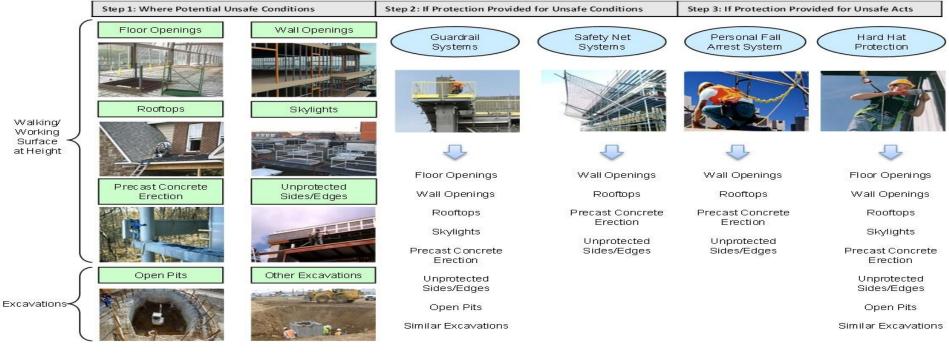
Applications: Smart Home for Elder Care

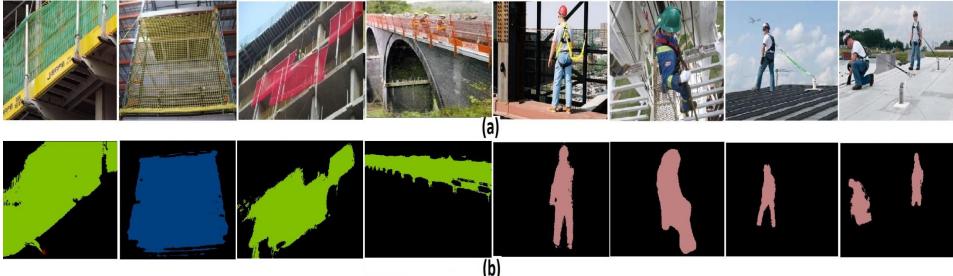


Two-stream CNNs

IEEE Trans. on T-IFS, vol.13, no.2, 2017 IEEE Trans. on T-IFS, vol.14, no.1, 2018 IEEE J. Biomedical and Health Informatics, 2014, 2015

Applications: Construction Safety





Predictions of High-Level Image Concepts

Ideas from Chinese Characters

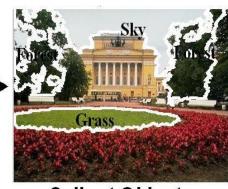
Automatic Salient

Object Detection

Human: animal can stand to walk



Original Image



Salient Objects



Classification



Semantic Image Concept: Garden

Predictions of High-Level Image Concepts

Ideas from Chinese Characters

Automatic Salient

Object Detection





Follow up: one person follows another



Original Image



Salient Objects



Classification



Semantic Image Concept: Garden

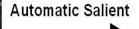
Predictions of High-Level Image Concepts

Ideas from Chinese Characters

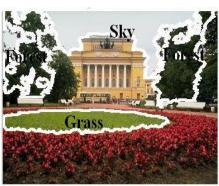




Original Image



Object Detection



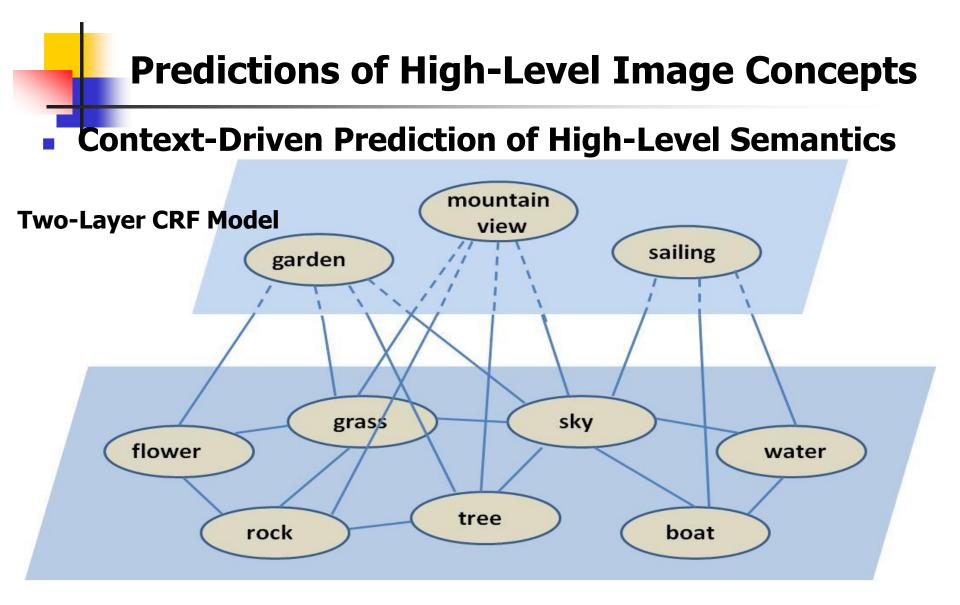
Salient Objects

Semantic Image

Classification



Semantic Image Concept: Garden



7. Conclusions

- Deep mixture algorithm to integrate diverse outputs from multiple experts with different but overlapping task spaces to generate a mixture network with larger outputs!
- Ontology-driven task group generation & identifying the interrelated learning tasks;
- Deep multi-task learning to exploit inter-class visual similarities and enhance their seperability;
- Deep Boosting to train the deep networks for the hard and easy object classes sequentially in an easy-to-hard way;
- Deep Collaborative Learning to learn multiple networks simultaneously & enhance each other;
- Knowledge Distillation for model compression for mobile usages.

Future Work: Human-like machine learning

- Developing Human-Like Learning Techniques
 ---directly learn from large-scale dirty data
 - Noise-free machine learning: machines should know which sources can be trusted more! –MIL?
 - Integrating decisions from different sources with different quality levels: machines should know which sources are more important and reliable!
 - Human Assistants: where, how & what? mixture intelligence?
 - Chemical Actions in Machine Learning
 - Multi-modal decision or information fusion