

دانشگاه تربیت دبیر شهید رجایی
دانشکده مهندسی کامپیوتر

برچسب زنی مبتنی بر جستجوی تصاویر با استفاده از مدل های عمیق

دانشجو: محیا محمدی کاشانی

استاد راهنما: دکتر سید حمید امیری

استاد مشاور: مهندس محمد مهدی حسینی

اسفند ماه ۱۳۹۷

House , Waistcoat



terrain



coral, reefs

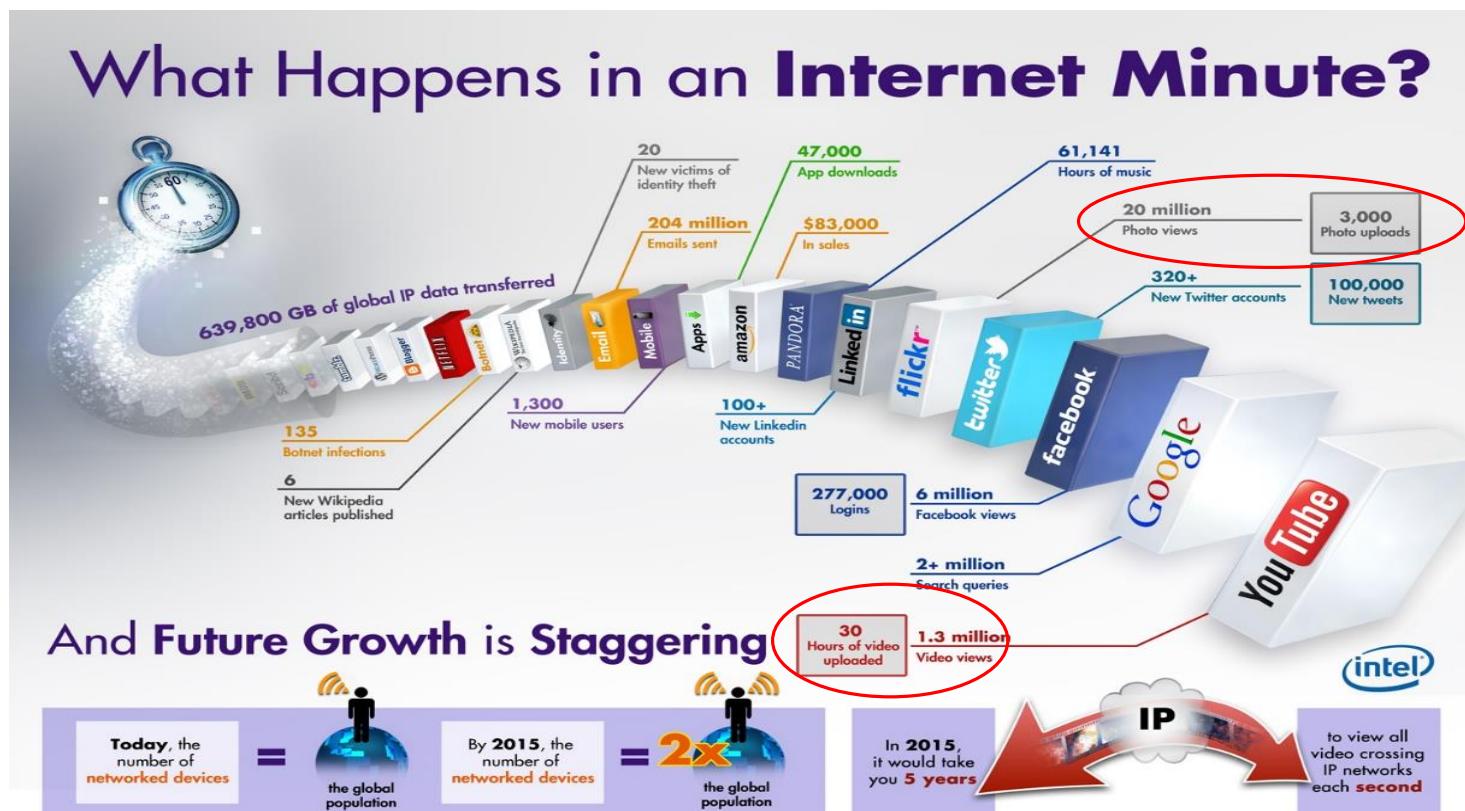


Girl, pullover

People



مقدمه



“What Happens in an Internet Minute”, By Lena Long, Dec 13th, 2013



سامانه برچسب زنی خودکار تصاویر

تصویر ورودی



تصویر خروجی

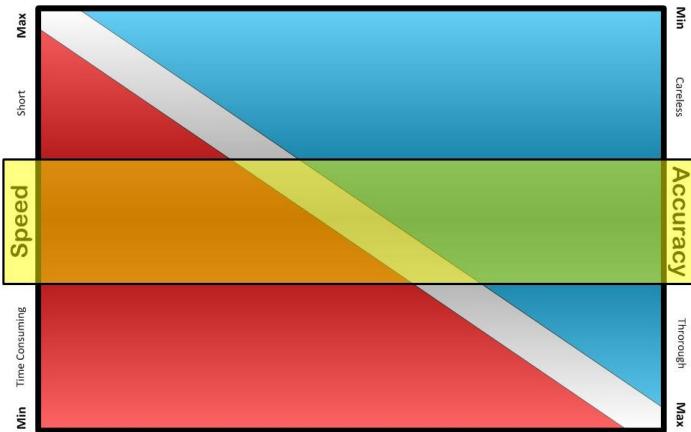
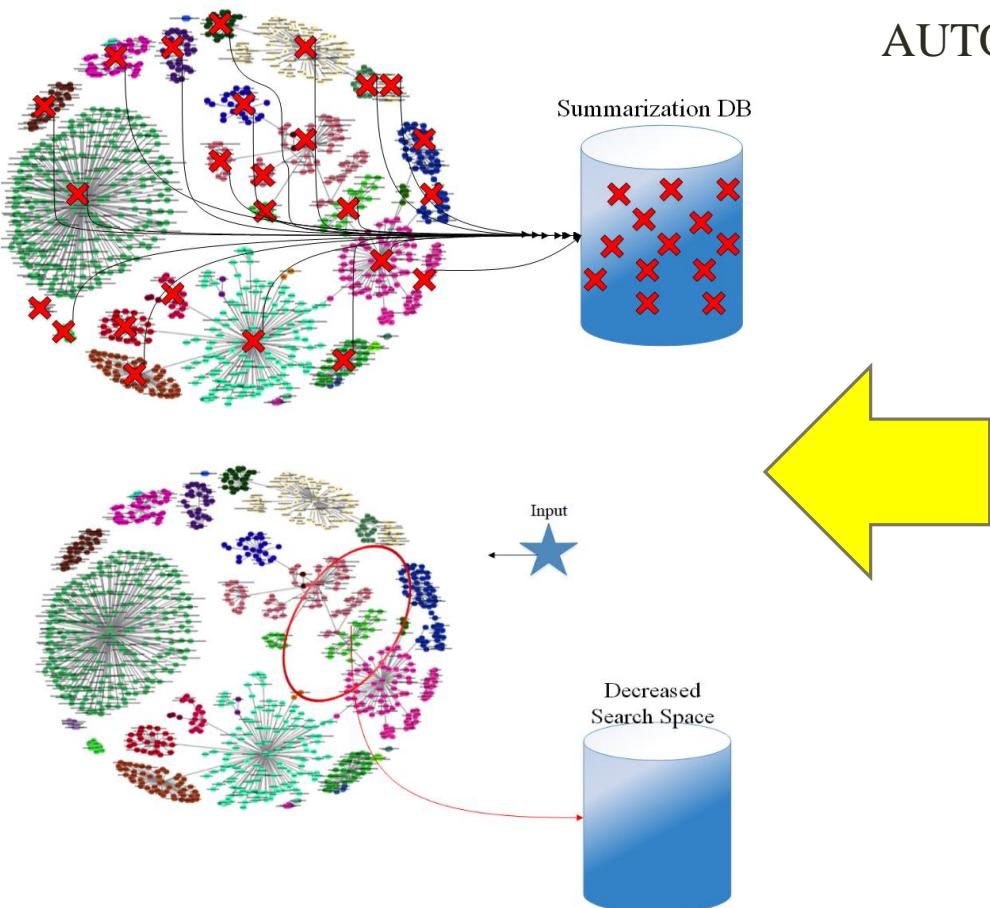


Panda, Black, Tree,
Wildlife



هدف تحقیق

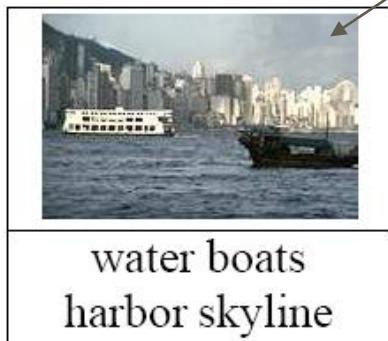
AUTOMATIC IMAGE ANNOTATOR



پیدا کردن نقطه بهینه برای کارآمدی عملکرد و سرعت
اجرای سامانه
“<http://bhexperience.blogspot.com>”



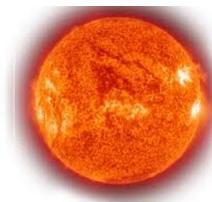
Sky



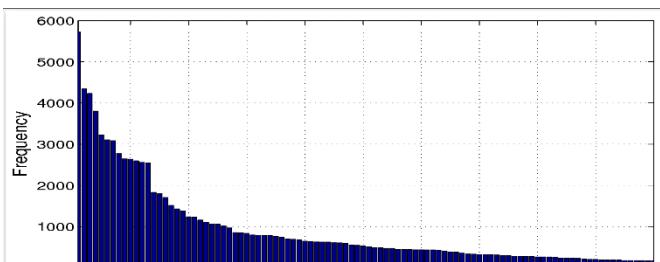
چالش برچسب زنی ناقص



شکاف معنایی



ناتوانی در به کارگیری الگوریتم ها برای داده های بزرگ



عدم توازن کلاس



cars, tracks, prototype



grass, flowers, petals

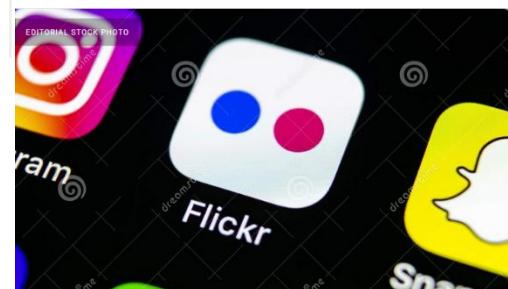


sky, grass, plane, lion



tree, grass, tiger, park

ابهام برچسب ها

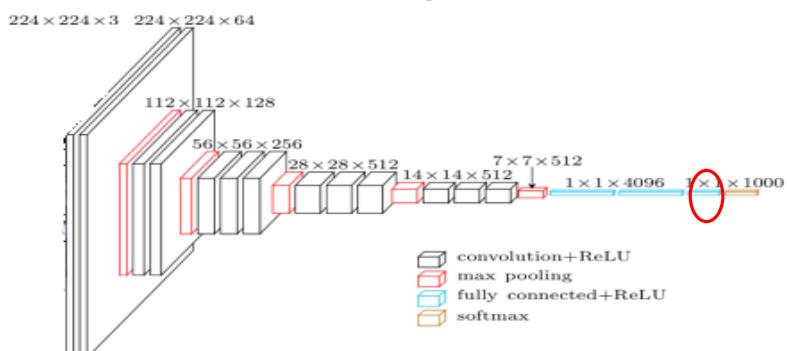


“ICT Pulse, www.dreamstime.com”



استخراج ویژگی با استفاده از مدل های یادگیری عمیق

معماری شبکه عصبی VGG-Net

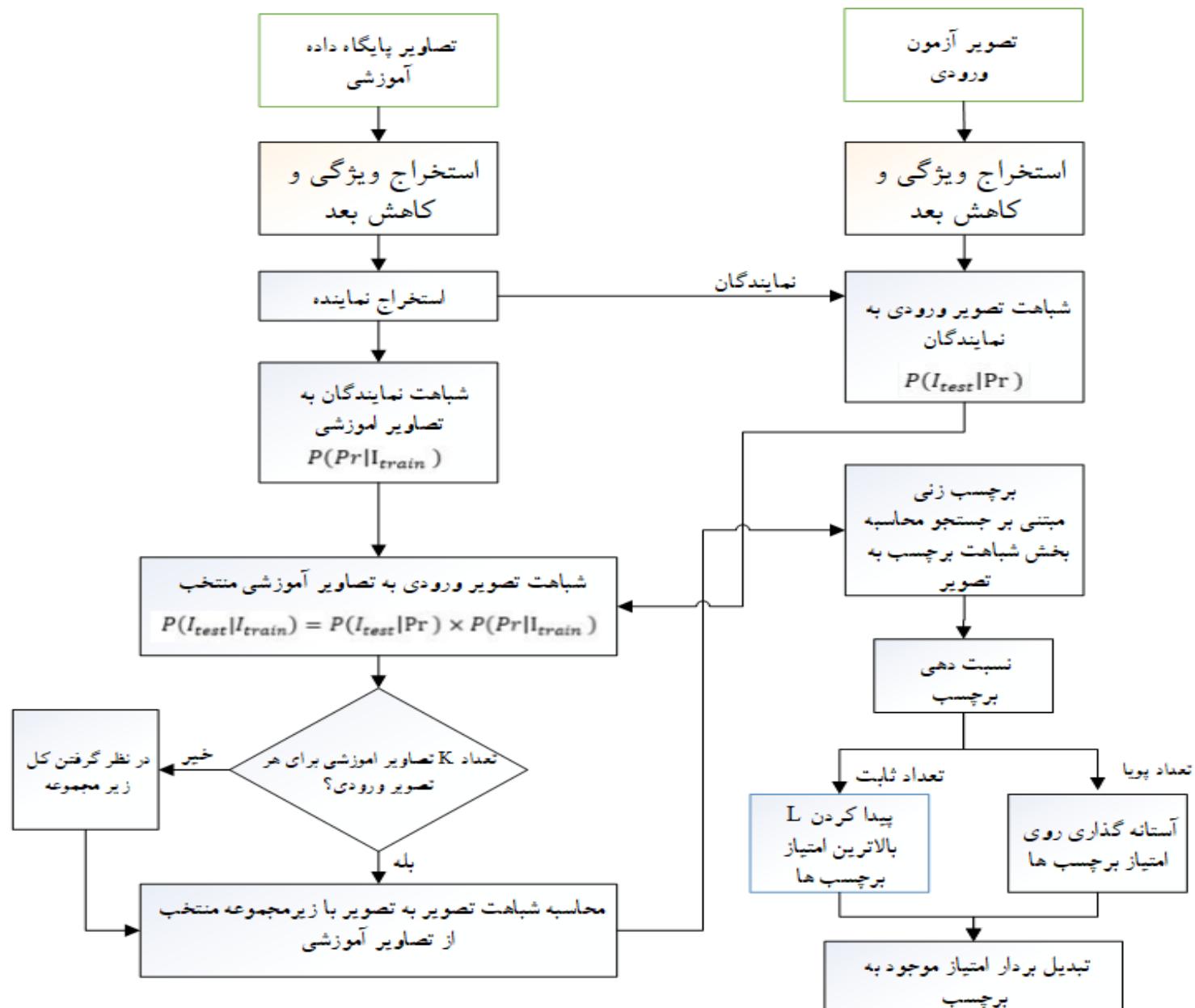


VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE SCALE IMAGE RECOGNITION, (K. Simonyan, 2015) (8)
شکل (8)

Other Frame Works

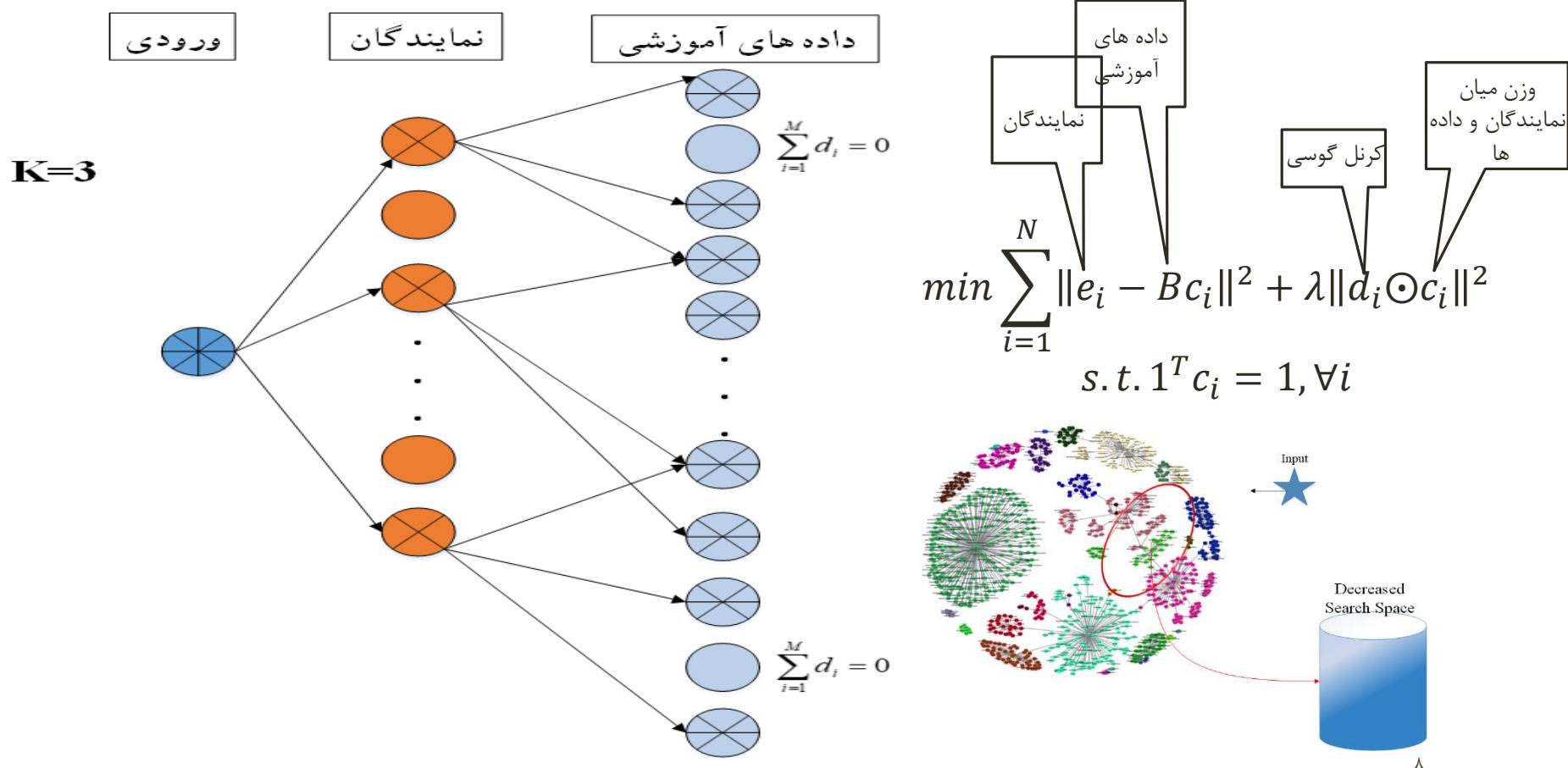
- ResNet101
- Dense169

بهترین قطعه بندی تصویر برای
استخراج ویژگی از مدل های عمیق





محاسبه بازسازی تنک ($P(I_{test} | Pr)$ و $P(Pr | I_{train})$)





روش پیشنهادی برای برچسب زنی تصاویر

$$P(I_{test}|I_{train}) = P(I_{test}|\text{Pr}) \times P(\text{Pr}|I_{train})$$

$$p(l_i|A) = \frac{p(A|l_i)p(l_i)}{p(A)}$$

مدل سازی مسئله به
پیدا کردن احتمال
ثانویه

$$y^* = \arg \max_i p(l_i|J)$$

بهترین برچسب برای تصویر
ورودی

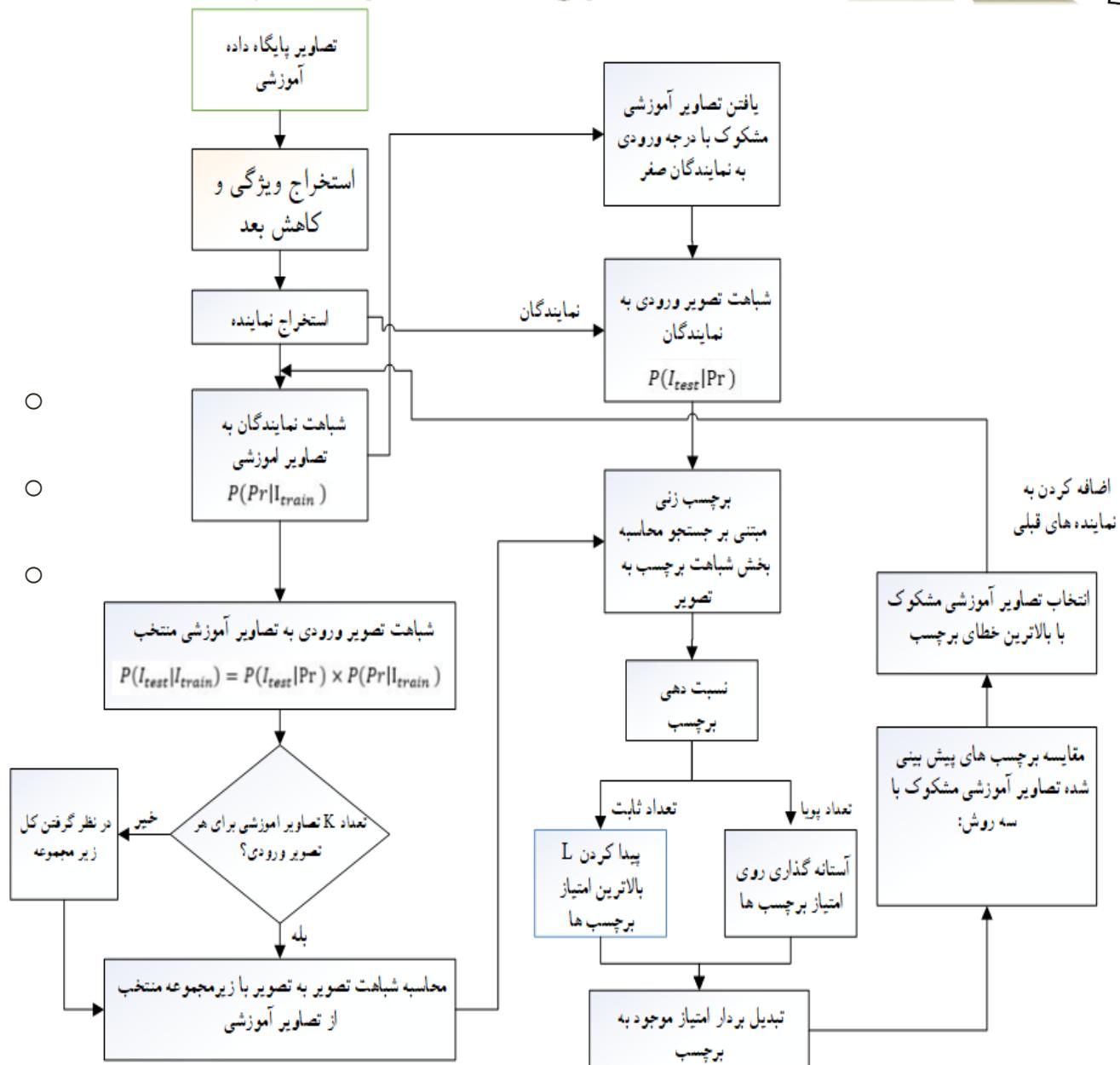
نسبت دهی برچسب

➤ نسبت دهی ثابت برچسب

➤ نسبت دهی پویا برچسب ← آستانه گذاری تطبیقی



- خطای مطلق برچسب‌های پیش‌بینی شده (RTIAFA)
- خطای نسبی برچسب‌های پیش‌بینی شده (RTIRFA)
- جمع وزن‌دار خطای مطلق برچسب‌های پیش‌بینی شده (RTIRA)





مجموعه داده های آزمایش شده

اطلاعات آماری مجموعه دادگان بر چسب زنی

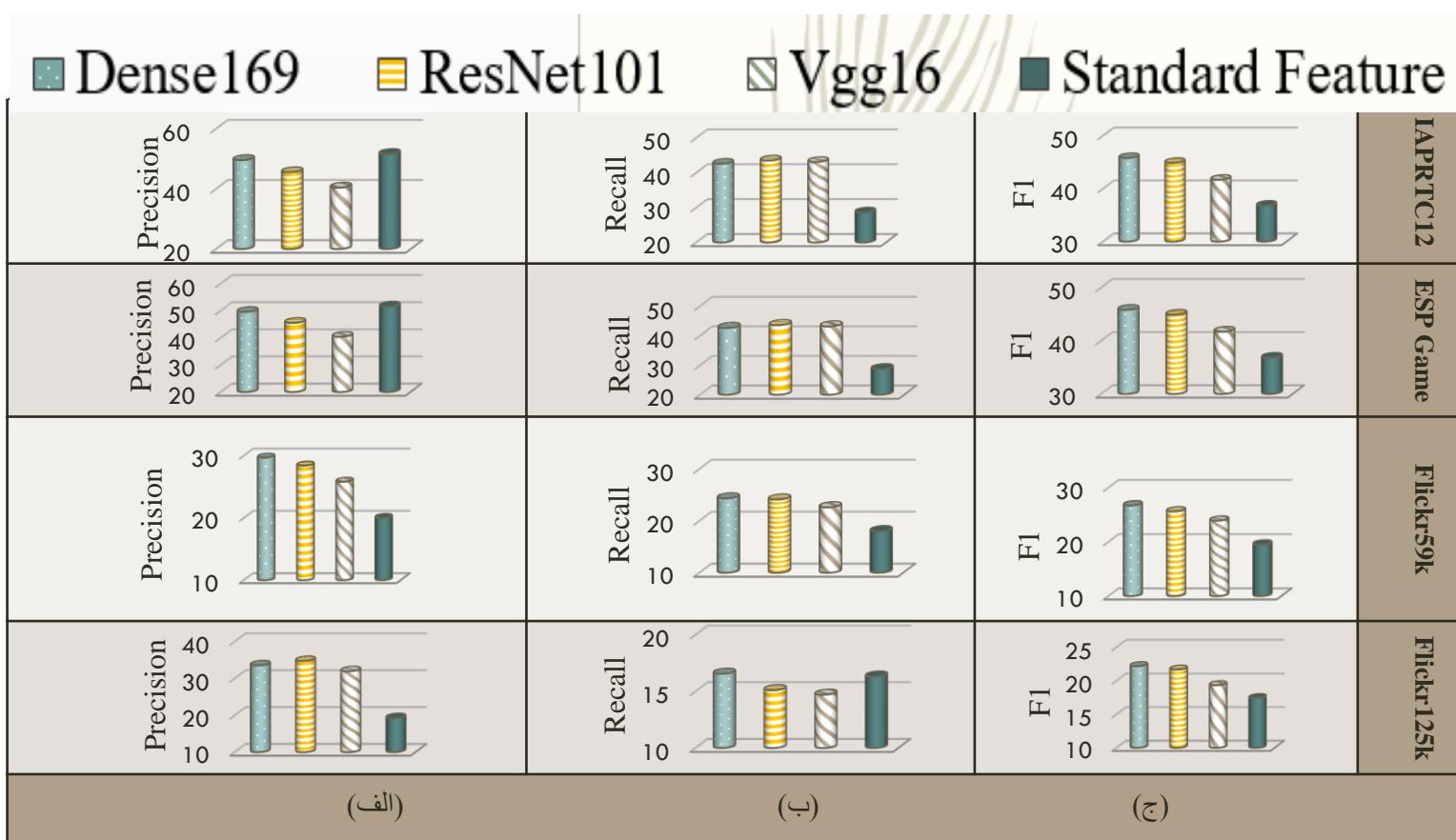
پایگاه داده	تعداد تصاویر	تعداد بر چسب ها	تعداد تصاویر آموزشی	تعداد تصاویر تست	تعداد بر چسب به ازای هر تصویر (حداکثر، میانه، میانگین)
ESP-Game	۲۰۷۷۰	۲۶۸	۱۸۶۸۹	۲۰۸۱	۴/۷, ۵, ۱۵
IAPR TC-12	۱۹۶۲۷	۲۹۱	۱۷۶۶۵	۱۹۶۲	۵/۷, ۵, ۲۳
Flickr59k	۵۹۰۸۳	۲۹۵	۴۱۳۵۹	۱۷۷۲۴	۵/۲۷, ۴, ۴۰
Flickr125k	۱۲۴۸۴۰	۵۶۸	۸۷۳۸۸	۳۷۴۵۲	۵/۸۸, ۵, ۶۴

Flickr tag set	
	houstontx, houston, downtown, harriscounty, southside, hdr, landscape , cityscape, ben, texas, usa, skyscrapers, canon40d, 40d
	350d, dacha, houseslippers, landscape , sigma1770f2845, uyma, slippers, flickr, explore, interestingness
	stone, stones, river, water, reflection, sky, blue, brown, wet, dry, underwater, ripples, song, itstonedme, vanmorrison, betws, bridge , amman, fotocyfer, sonyalpha350
	bridge , motel, seattle, sign, todbates

IAPR TC-12 images		
Ground-truth	cloud, eagle, sky, flower, rock, sky, mountain, rock	car, cloud, grass, house, sky, tree



مقایسه و بررسی ویژگی های یادگیری ژرف مختلف

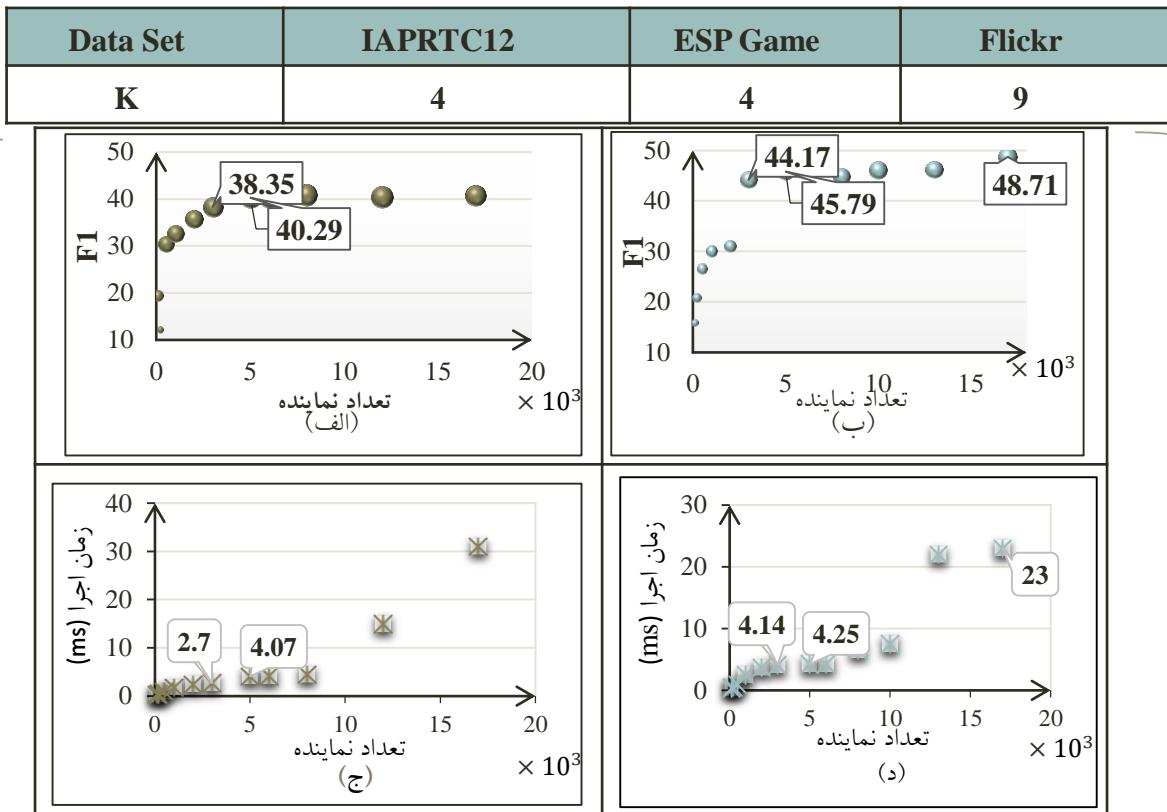


ارزیابی سامانه‌ی برچسب‌زنی کاهش فضای جستجو به ازای ویژگی‌های مدل‌های یادگیری عمیق بر حسب میانگین دقت (الف)، میانگین یادآوری (ب) و F1 (ج)



آنالیز حساسیت تعداد نمایندگان

جدول ۱) تعیین پارامتر K نزدیکترین همسایگی طی بازسازی وزن ها بر روی مجموعه دادگان مختلف

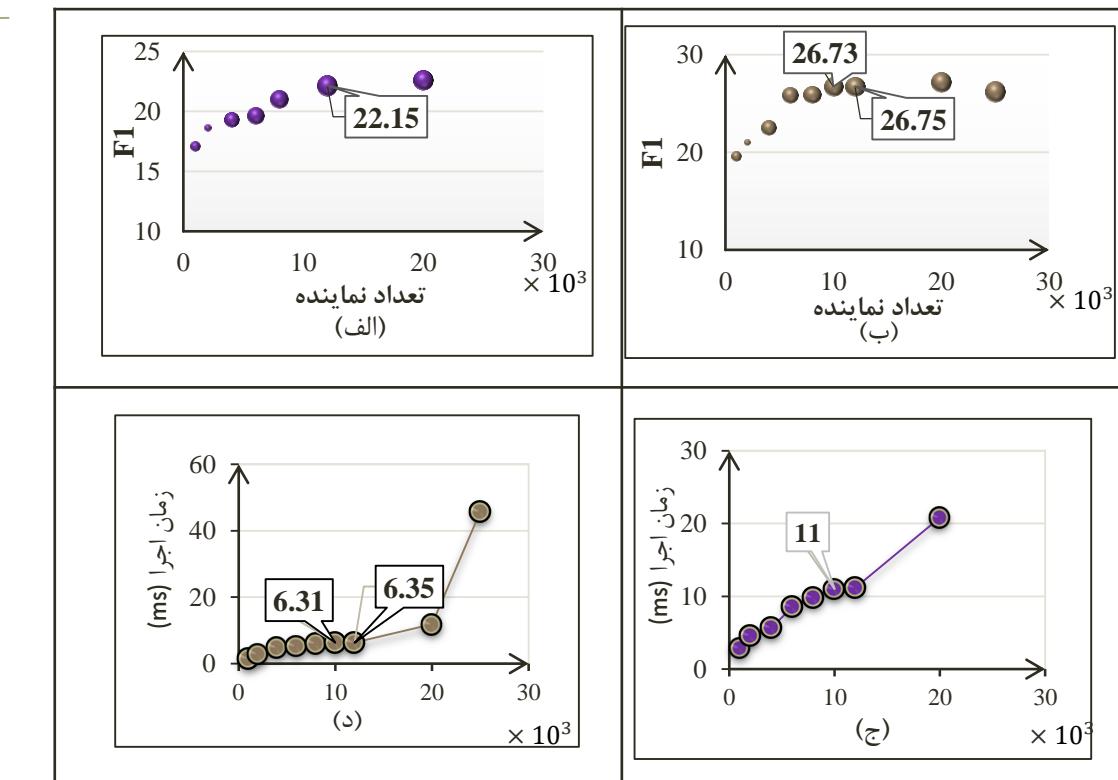


ارزیابی کارایی سامانه برچسبزنی به ازای نمایندگان و زمان اجرای برچسبزنی برای یک تصویر ورودی



آنالیز حساسیت تعداد نمایندگان

Flickr125k



Flickr59k

ارزیابی کارایی سامانه برچسبزنی به ازای نمایندگان و زمان اجرای برچسبزنی برای یک تصویر ورودی

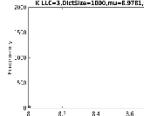
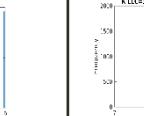
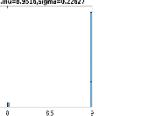
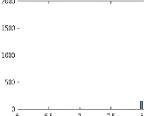
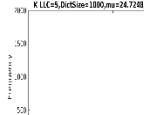
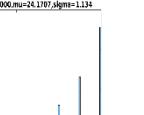
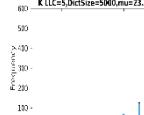
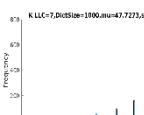
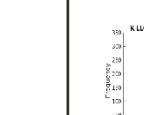
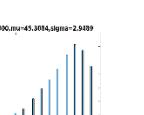
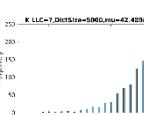
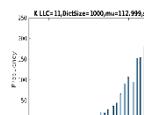
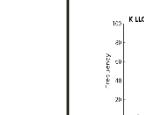
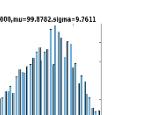
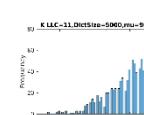


آنالیز حساسیت بر روی پارامتر K

K	IAPRTC12	Esp-Game	Flicker59k	Flicker125k
3	41/8	35/2	21/9	18/7
5	44/18	40/29	22/81	19/54
7	45/79	40/44	24/56	21/45
11	46/23	40/89	26/73	22/15



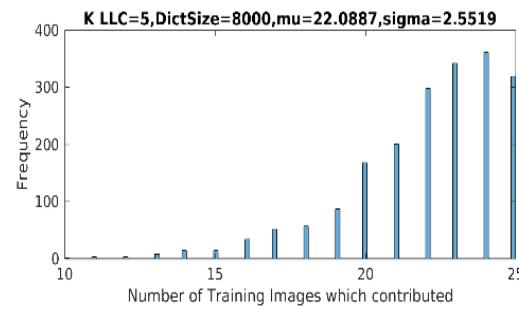
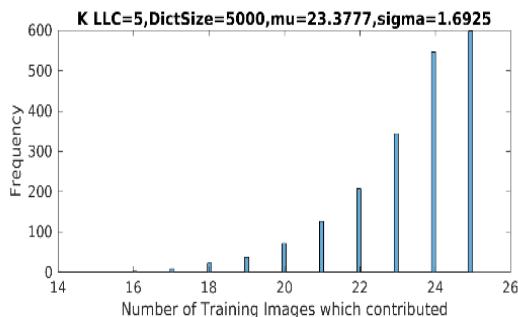
تحلیل تصاویر آموزشی موثر به ازای هر تصویر ورودی

M K	1K	3K	5K	8K
3	 <p>K LLC=3, DictSize=1000, mu=4.9781, sigma=0.14843</p> <p>0/14 δ و 7/8 μ</p>	 <p>K LLC=3, DictSize=5000, mu=8.9515, sigma=0.22927</p> <p>0/22 δ و 8/95 μ</p>	 <p>K LLC=3, DictSize=5000, mu=8.9601, sigma=0.34129</p> <p>0/34 δ و 9/8 μ</p>	 <p>K LLC=3, DictSize=8000, mu=8.7054, sigma=0.66585</p> <p>0/66 δ و 7/8 μ</p>
5	 <p>K LLC=5, DictSize=1000, mu=24.7289, sigma=0.57545</p> <p>0/57 δ و 24/72 μ</p>	 <p>K LLC=5, DictSize=3000, mu=24.1107, sigma=1.134</p> <p>1/31 δ و 24/17 μ</p>	 <p>K LLC=5, DictSize=5000, mu=23.3777, sigma=1.5925</p> <p>1/69 δ و 23/72 μ</p>	 <p>K LLC=5, DictSize=8000, mu=22.8897, sigma=2.5539</p> <p>2/55 δ و 22/08 μ</p>
7	 <p>K LLC=7, DictSize=1000, mu=42.7273, sigma=1.4227</p> <p>1/42 δ و 47/72 μ</p>	 <p>K LLC=7, DictSize=3000, mu=45.3084, sigma=2.9409</p> <p>2/95 δ و 45/3 μ</p>	 <p>K LLC=7, DictSize=5000, mu=42.4038, sigma=4.2682</p> <p>2/4 δ و 42/48 μ</p>	 <p>K LLC=7, DictSize=8000, mu=39.3645, sigma=5.756</p> <p>5/76 δ و 39/37 μ</p>
11	 <p>K LLC=11, DictSize=1000, mu=111.9999, sigma=5.1457</p> <p>1/5 δ و 112 μ</p>	 <p>K LLC=11, DictSize=3000, mu=99.8702, sigma=9.7611</p> <p>9/76 δ و 87/99 μ</p>	 <p>K LLC=11, DictSize=5000, mu=90.9292, sigma=13.2781</p> <p>13/28 δ و 92/90 μ</p>	 <p>K LLC=11, DictSize=8000, mu=81.053, sigma=18.5526</p> <p>16/53 δ و 83/05 μ</p>

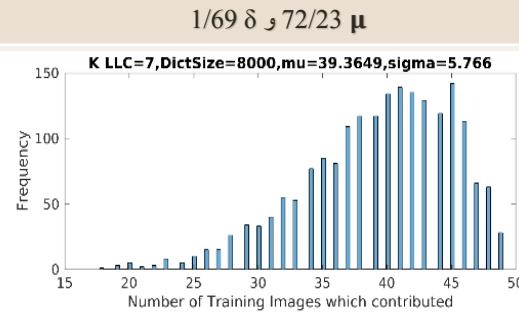
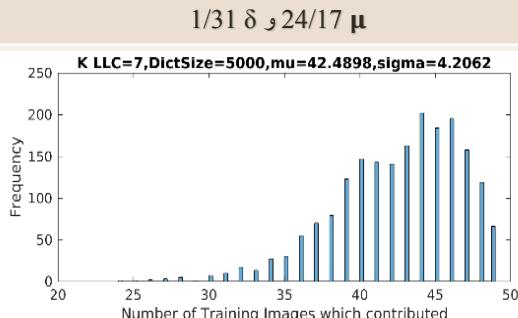


5K

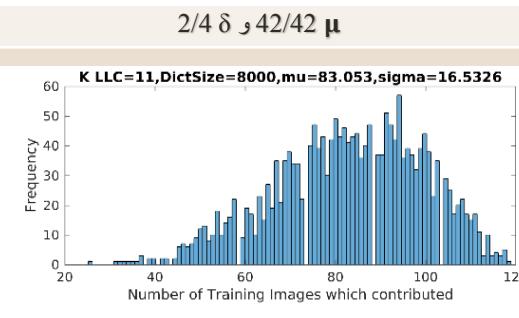
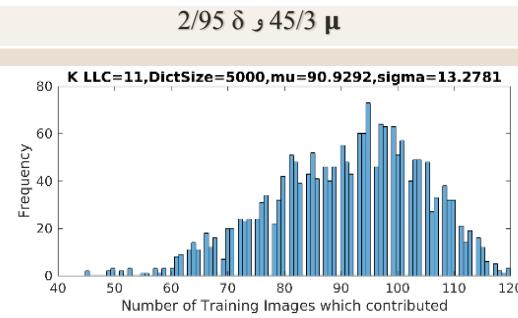
8K

M
K

5



7



11

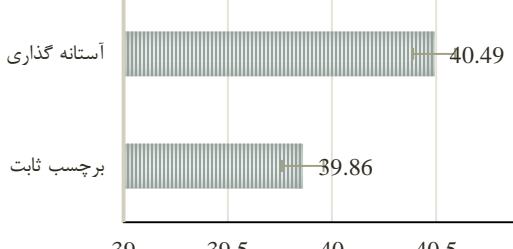


نسبت دهی برچسب ثابت و پویا (تطبیقی)

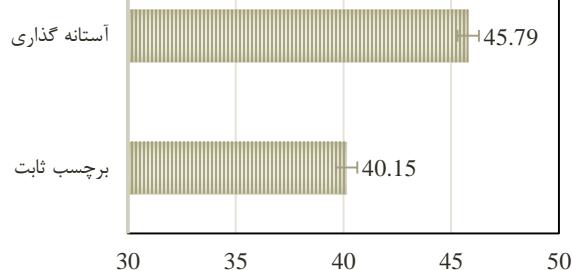
ESP Game Flickrl25k

IAPRTC12 Flickr59k

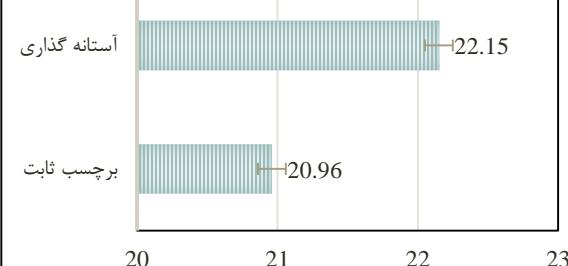
(ب)



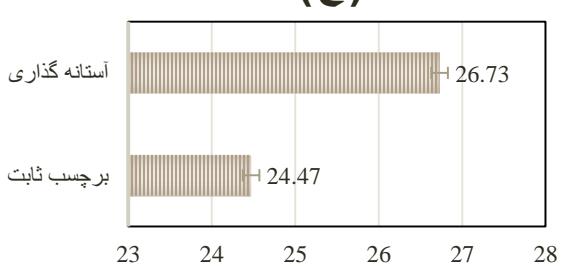
(الف)



(د)



(ج)



شکل (۱۵) مقایسه کارایی معیار F_1 به ازای دو رویکرد برچسبزنی تصویر



Method	Feature	ESP Game			IAPRTC-12			Flickr59k			Flickr125k		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
JEC	Standard	0.22	0.25	0.23	0.28	0.29	0.29	-	-	-	-	-	-
MBRM	Standard	0.18	0.19	0.19	0.24	0.23	0.24	-	-	-	-	-	-
Tag Prop (ML)	Standard	0.49	0.2	0.28	0.48	0.25	0.33	-	-	-	-	-	-
Tag Prop (δ ML)	Standard	0.39	0.27	0.32	0.46	0.35	0.4	-	-	-	-	-	-
Fast Tag	Standard	0.46	0.22	0.3	0.47	0.26	0.34	-	-	-	-	-	-
KSVM-VT	Standard	0.33	0.32	0.33	0.47	0.29	0.36	-	-	-	-	-	-
NMF-KNN	Standard	0.33	0.26	0.29	-	-	-	-	-	-	-	-	-
CCA-KNN	CNN-VGG16	0.46	0.36	0.41	0.45	0.38	0.41	-	-	-	-	-	-
2PKNN	Standard	0.51	0.23	0.32	0.49	0.32	0.39	-	-	-	-	-	-
2PKNN(ML)	Standard	0.53	0.27	0.36	0.54	0.37	0.44	-	-	-	-	-	-
MVSAE	SAE	0.47	0.28	0.34	0.43	0.38	0.4	-	-	-	-	-	-
Mvg-NMF	Standard	0.41	0.33	0.37	0.46	0.4	0.43	-	-	-	-	-	-
2PKNN-without Normalization	CNN-Dense169	0.49	0.32	0.39	0.53	0.32	0.4	0.34	0.25	0.29	0.42	0.18	0.25
2PKNN	CNN-Dense169	0.44	0.4	0.42	0.50	0.41	0.44	0.31	0.29	0.30	0.33	0.23	0.27
DSSM (our method)	CNN-Dense169	0.44	0.37	0.405	0.49	0.42	0.44	0.3	0.25	0.27	0.33	0.16	0.22
RTIAFA (our method)	CNN-Dense169	0.43	0.39	0.41	0.45	0.43	0.43	0.30	0.26	0.27	0.34	0.19	0.24
RTIRFA (our method)	CNN-Dense169	0.439	0.376	0.40	0.45	0.44	0.45	0.25	0.28	0.27	0.31	0.18	0.23
RTIRA (our method)	CNN-Dense169	0.44	0.37	0.403	0.46	0.44	0.45	0.27	0.26	0.27	0.32	0.18	0.22



پیش بینی بر چسب بر روی تصاویر مجموعه داده ها



Predicted tags:
Mountain,
river, tree

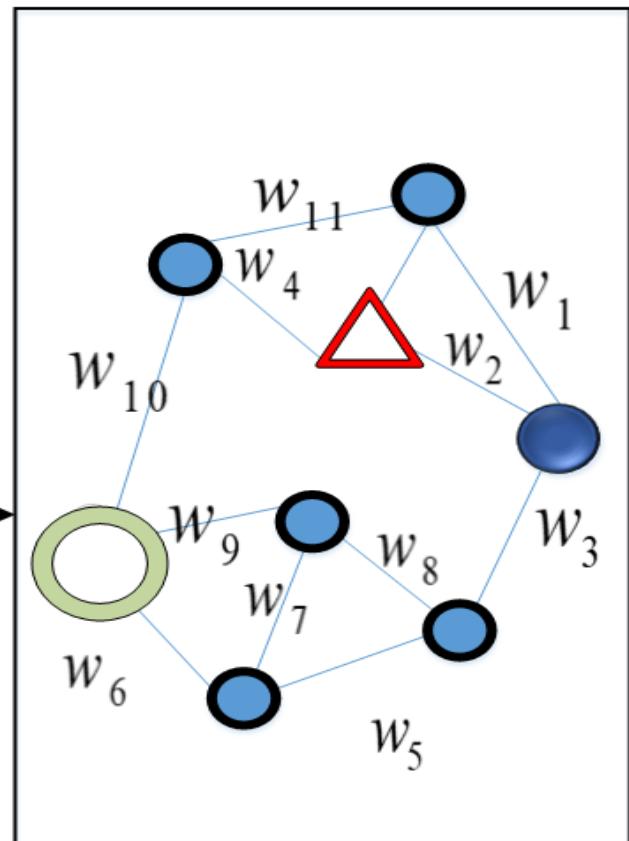
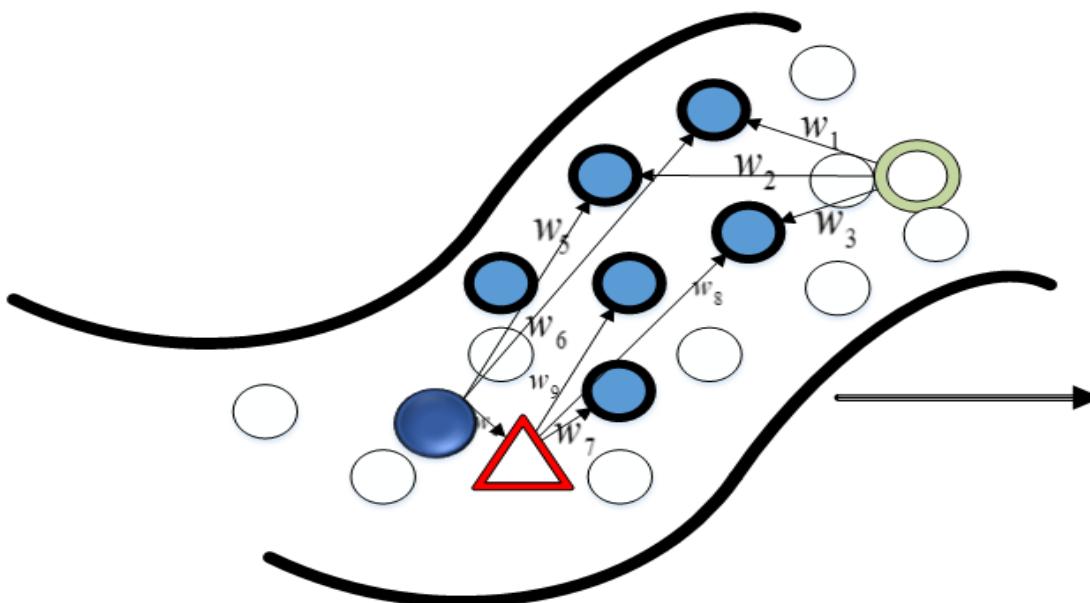
Predicted tags:
Boy, child, girl,
man, wall,
woman

Predicted tags:
Portrait, canon,
Nikon, animal,
dog, cute

Predicted tags:
Wildlife,
animal, nature,
photos



خلاصه سازی پایگاه داده آموزشی برای برچسبزنی خودکار تصاویر مبتنی بر جستجو



نمونه گراف ساخته شده از نمایندگان و مجموعه آموزشی



انتشار چند برچسبی روی گراف

$$Q(F) = \sum_{i=M, i \notin N}^P \sum_{j: x_j \in \mathcal{N}(x_i)} W_{ij} \|F_i - F_j\|^2 + \gamma \sum_{i=M, i \notin N}^P \|F_i - Y_i\|^2 \quad (1)$$

الگوریتم خلاصه‌سازی پایگاه داده

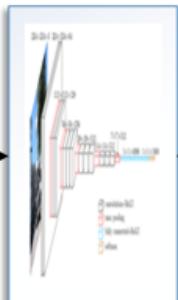
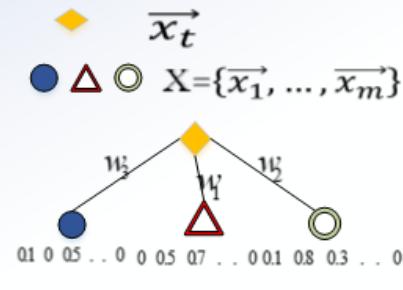
- ✓ وزن گراف برای رئوس نمایندگان و داده‌های آموزشی از طریق بازسازی تقریبی کدگذاری خطی محدودیت محلی (LLC)
- $W > 0$ ✓
- $W = \frac{W^T + W}{2}$ ✓
- $0 < \alpha < 1$ ، $F(t+1) = \alpha F(t) + (1-\alpha)Y$ ✓
- $F^{(t+1)}(u)_{u \in X} = F^{(t)}(u)_{u \in X}$ ✓



روش پیشنهادی سامانه برحسب زنی با خلاصه سازی پایگاه داده

فاز آزمون روش خلاصه سازی پایگاه داده

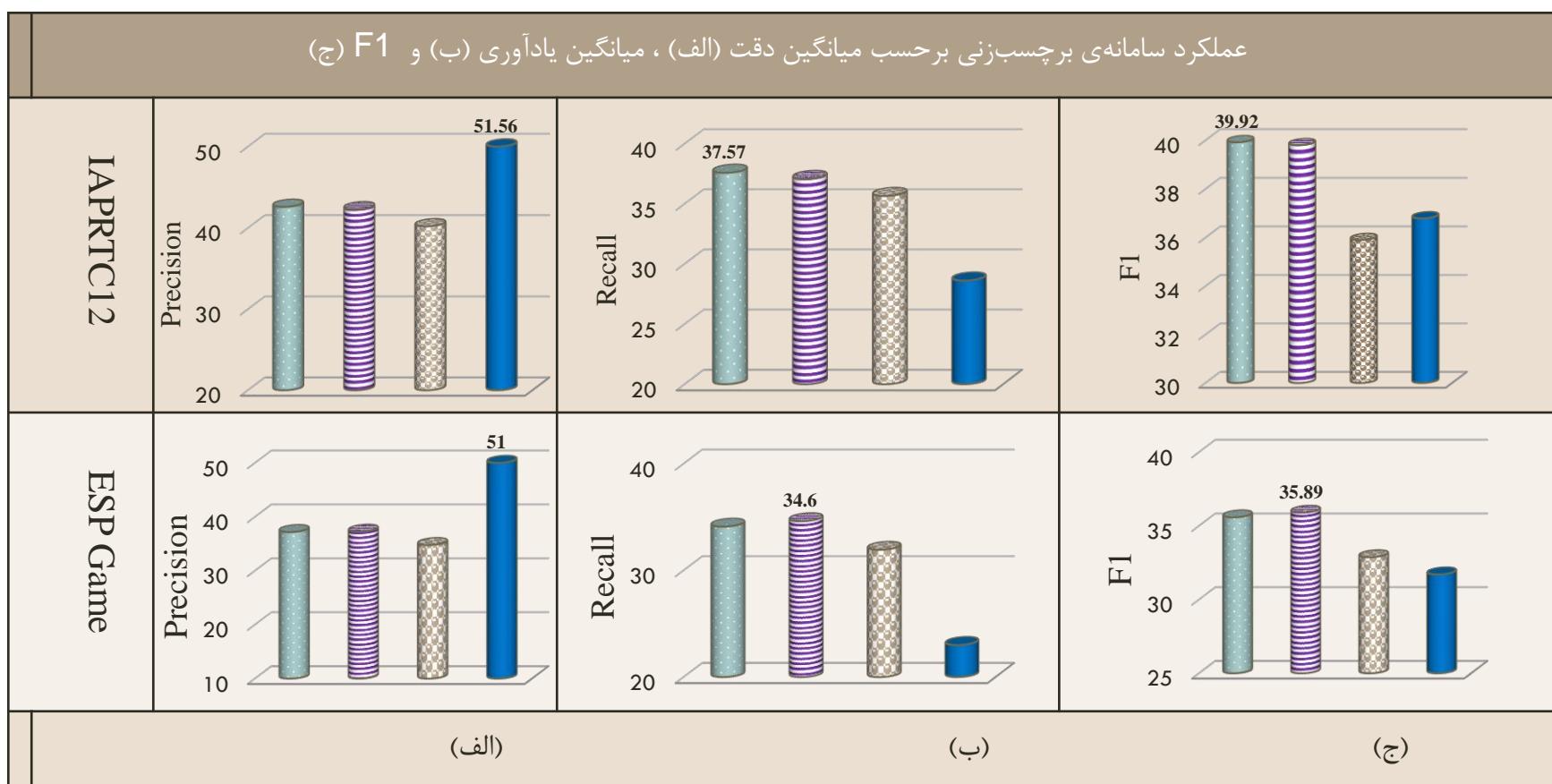
تصویر ورودی

استخراج
ویژگیپیدا کردن K نزدیک ترین همسایه
میان نمایندگانمحاسبه امتیاز
برحسب برای
تصویر ورودی

$$\sum w_{ij} p_j$$
آستانه
گذاری روی
برحسب هاBear, River,
Grass, Brown



مقایسه عملکرد سامانه‌ی برچسبزنی بر روی ویژگی‌های مختلف نسبت به روش پایه

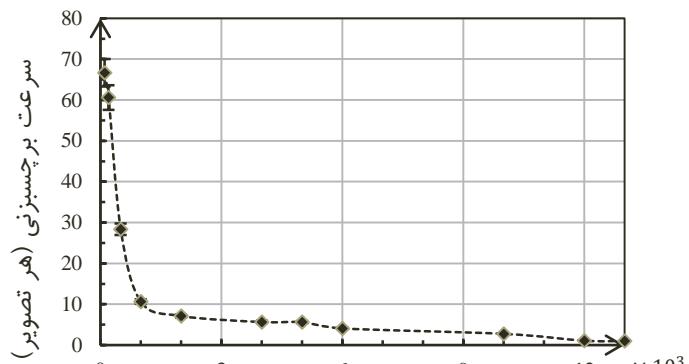




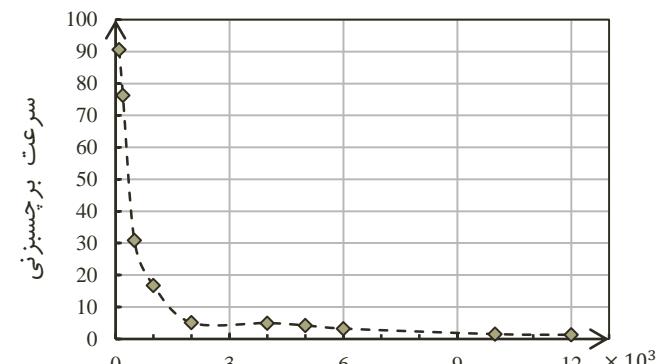
آنالیز حساسیت بر روی تعداد نمایندگان

IAPRTC12

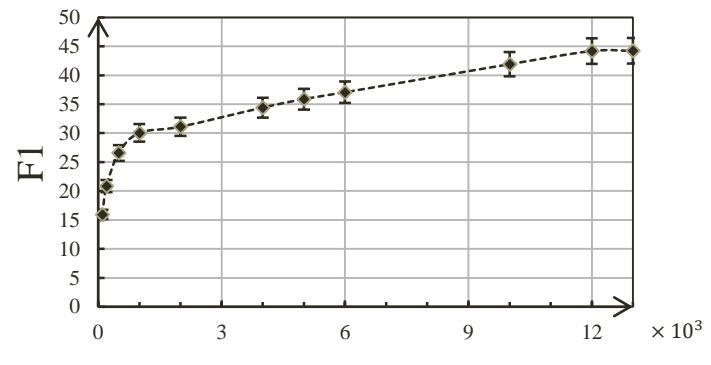
ESP Game



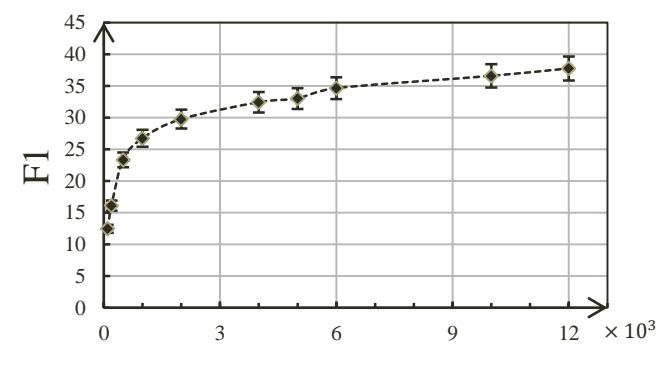
(ج)



(الف)



(د)

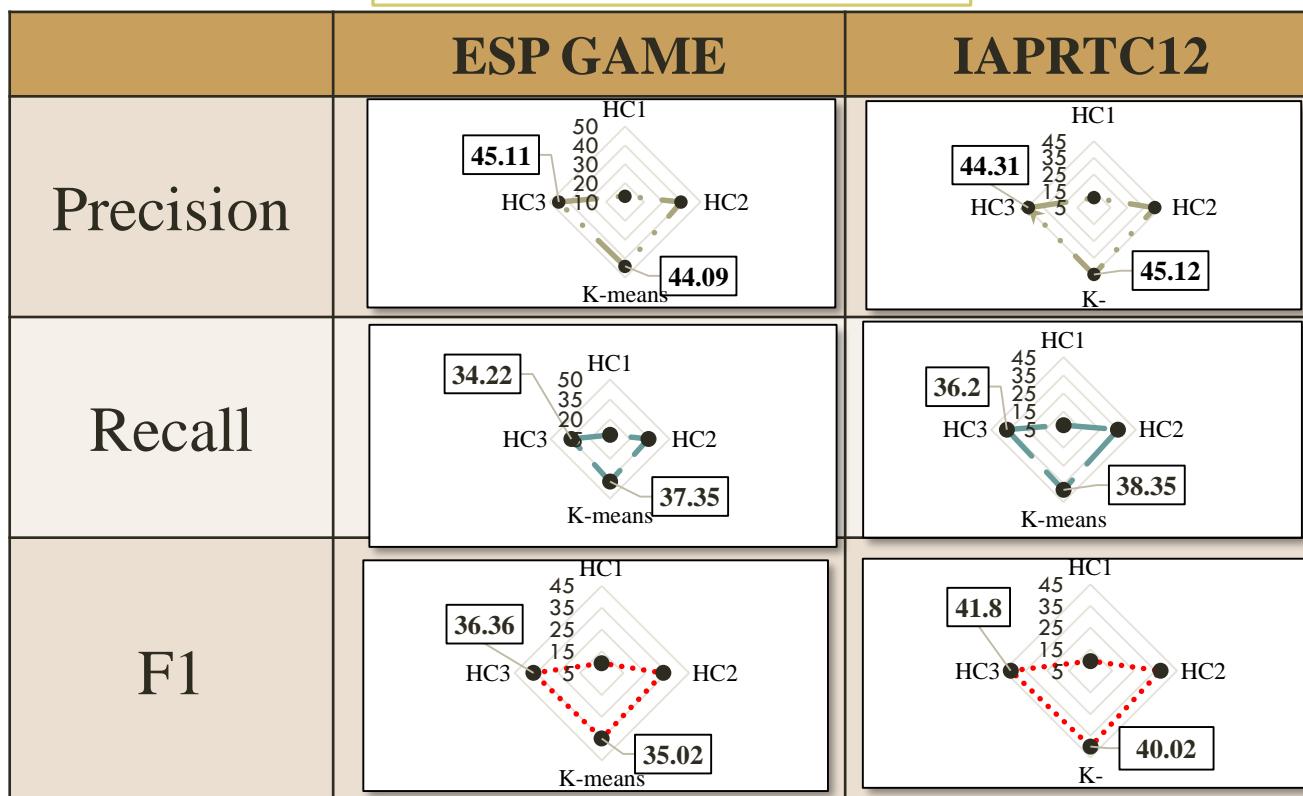


(ب)



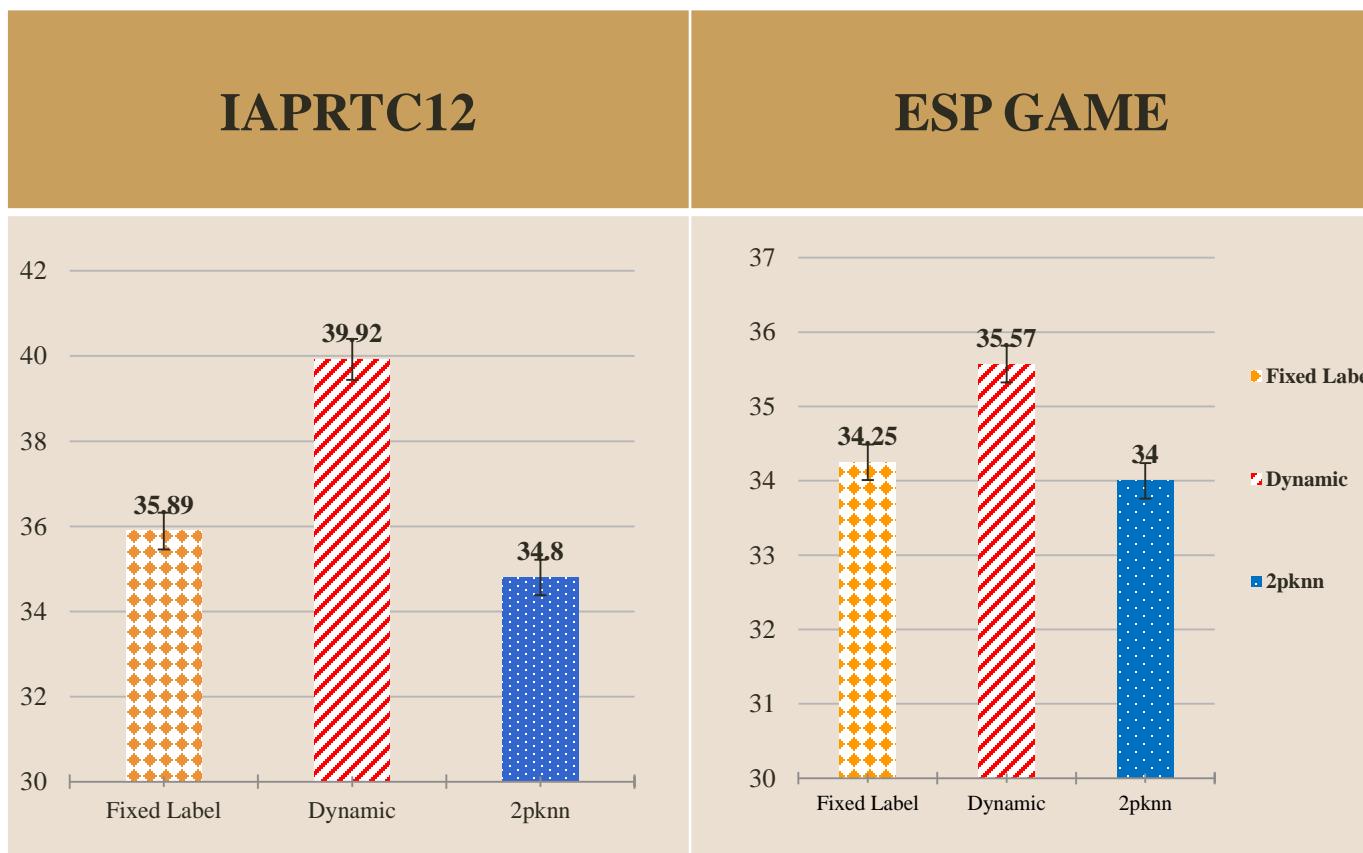
مقایسه دو روش استخراج نماینده

HC1: hierarchical clustering with 1 layer
 HC2 : hierarchical clustering with 2 layer
 HC3 : hierarchical clustering with 3 layer





مقایسه روش های نسبت دهی بر چسب پویا و ثابت





مقایسه روش پیشنهادی بر چسب زنی با خلاصه سازی با سایر روش های بر چسب زنی

Method	Feature	ESP Game			IAPRTC-12		
		P	R	F1	P	R	F1
JEC	Standard	0.22	0.25	0.23	0.28	0.29	0.29
MBRM	Standard	0.18	0.19	0.19	0.24	0.23	0.24
Tag Prop (ML)	Standard	0.49	0.2	0.28	0.48	0.25	0.33
Tag Prop (δ ML)	Standard	0.39	0.27	0.32	0.46	0.35	0.4
2PKNN	Standard	0.51	0.23	0.32	0.49	0.32	0.39
2PKNN(ML)	Standard	0.53	0.27	0.35	0.54	0.37	0.44
Fast Tag	Standard	0.46	0.22	0.3	0.47	0.26	0.34
NMF-KNN	Standard	0.33	0.26	0.29	-	-	-
CCA-KNN	Standard	0.46	0.36	0.41	0.45	0.38	0.41
KSVM-VT	Standard	0.33	0.32	0.33	0.47	0.29	0.36
2PKNN-without Normalization	CNN(Dense169)	0.49	0.32	0.39	0.531	0.321	0.4
2PKNN	CNN(Dense169)	0.44	0.39	0.42	0.49	0.41	0.44
TDSNLP+2PKNN	CNN(Dense169)	0.44	0.28	0.342	0.45	0.34	0.39
TDSNLPFT	CNN(Dense169)	0.37	0.34	0.35	0.45	0.39	0.42
TDSNLPDT	CNN(Dense169)	0.44	0.37	0.36	0.47	0.42	0.44



آموزش نماینده معنایی حاشیه گسترش یافته برای برچسبزنی بلاذرنگ تصویر

- Coupled Dictionary Learning
- Optimization
 1. Sparse Representation
 2. Update Prototypes

تابع هدف مسئله

تعمیم پذیری

LMSPL • آموزش همزمان نماینده‌گان برای ماهیت‌های معنایی و بصری

عدم انسجام ← رویکرد حاشیه گسترش یافته

$$(D^I, D^L, \alpha) = \operatorname{argmin} \sum_{i=1}^N \|x^i - D^I \alpha^i\|_2^2 + \lambda \sum_{i=1}^N \sum_{t=1}^T l(y_t^i, d_t^L \alpha^i) \quad (1)$$

$$s.t \|\alpha^i\|_1 \leq 1, \alpha_k^i \geq 0, d_k^I \in \mathbb{D} = \{d \in \mathbb{R}^{(M)}, \|d\|_2^2 \leq 1\}, \forall k \in \{1, \dots, K\}$$



Algorithm 1: Coupled Prototype Learning

Input : The set of training images $\mathbf{X} = [x^1, \dots, x^N]$ and their corresponding label vectors $\mathbf{Y} = [y^1, \dots, y^N]$.
discriminative regularization λ .

The number of prototypes K .

Sparsity factor is considered 1 which is consistent with data normalization.

The number of repeats $R = 100$.

Output: The learned prototypes \mathbf{D}^I and \mathbf{D}^L .

Step 1: Data Normalization

$$x^i \leftarrow \frac{x^i}{\|x^i\|_2}, \forall i \in \{1, \dots, n\}$$

Step 2: Initialize visual and semantic prototypes

$\mathbf{D}^C \leftarrow$ Apply $k - means$ algorithm on \mathbf{Z} and provide K prototypes.

Step 3: Normalize visual prototypes

$$d_k^L \leftarrow \frac{d_k^L}{\|d_k^L\|_2}, \forall k \in \{1, \dots, K\}$$

$$d_k^I \leftarrow \frac{d_k^I}{\|d_k^I\|_2}, \forall k \in \{1, \dots, K\}$$

Step 4: Optimize prototypes

Solve Equation (4) by alternate updating of \mathbf{Q} and \mathbf{D}^C while the other is assumed to be fixed:

for $r \leftarrow 1$ to R do

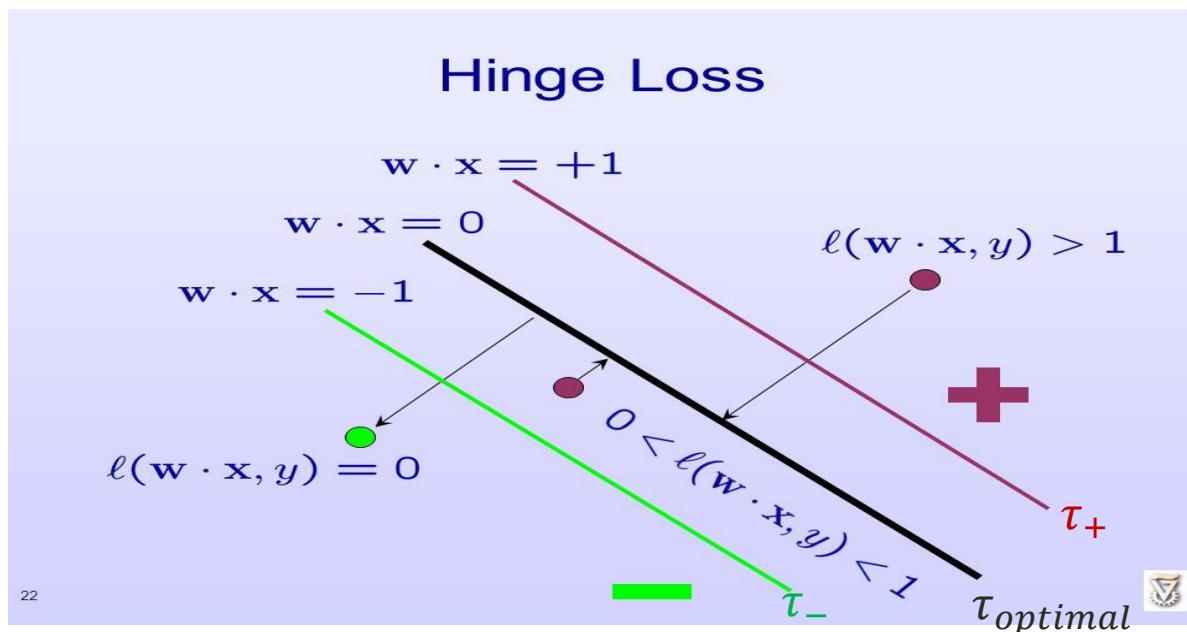
- Update sparse representations $\mathbf{Q} = [\alpha^1, \dots, \alpha^N]$ using Lasso technique by solving ($\forall i \in \{1, \dots, N\}$):

$$(\alpha^i) = argmin \left\| \begin{bmatrix} x^i \\ \lambda y^i \end{bmatrix} - \begin{bmatrix} \mathbf{D}^I \\ \lambda \mathbf{D}^L \end{bmatrix} \alpha^i \right\|_2^2 \text{ s.t. } \|\alpha^i\|_1 \leq 1, \alpha^i \geq 0.$$

- Update each column of $\mathbf{D}_{(r)}^C$ using Equation (3)



تنظیم دقیق نمایندگان معنایی برای امتیادهای حاشیه گسترش یافته



Apply Gradient Descent for
Fine-Tuning



Algorithm 2: Large Margin Fine-tune

Input : The set of normalized training images $\mathbf{X} = [x^1, \dots, x^N]$ and their corresponding label vectors $\mathbf{Y} = [y^1, \dots, y^N]$.

The learned prototypes \mathbf{D}^I and \mathbf{D}^L of Algorithm 1.

Regularization β for $l_2 - norm$.

The number of repeats $R = 100$.

Output: The final prototypes \mathbf{D}^I and \mathbf{D}^L .

Step 1: Obtain sparse representation for each data sample using image modality

for $i \leftarrow 1$ to N do

$$\alpha^i = argmin \|x^i - \mathbf{D}^I \alpha^i\|_2^2 \text{ s.t. } \|\alpha^i\|_1 \leq 1, \alpha^i \geq 0.$$

Step 2: Determine appropriate margins for positive and negative labels

Find $\tau_{optimal}$ (optimal threshold for annotation) which maximize $f - measure$ on training samples. Scoring function for each data sample is defined as:

$$Score(i, t) = \mathbf{d}_t^L \alpha^i, \forall i \in \{1, \dots, N\}, \forall t \in \{1, \dots, T\}.$$

Calculate τ_+ and τ_- using Equation (10).

Step 3: Gradient descent

for $m \leftarrow 1$ to M do

$$\text{Calculate } Score(i, t) = \mathbf{d}_t^L \alpha^i, \forall i \in \{1, \dots, N\}, \forall t \in \{1, \dots, T\}.$$

Determine pairs of (i, t) for which the loss function defined in Equation (2) is not zero.

Calculate partial derivatives of $\mathbf{D}_{t,k}^L$, $\forall t \in \{1, \dots, T\}$,

$\forall k \in \{1, \dots, K\}$ using Equation (3).

$$\mu \leftarrow 1.$$

do

$$\mathbf{D}^{L'} \leftarrow \mathbf{D}^L + \mu \nabla_{\mathbf{D}^L}^f$$

$$\mu \leftarrow \frac{\mu}{2}.$$

while $f(\mathbf{D}^{L'}) > f(\mathbf{D}^L)$ or $\mu > 2^{-20}$;

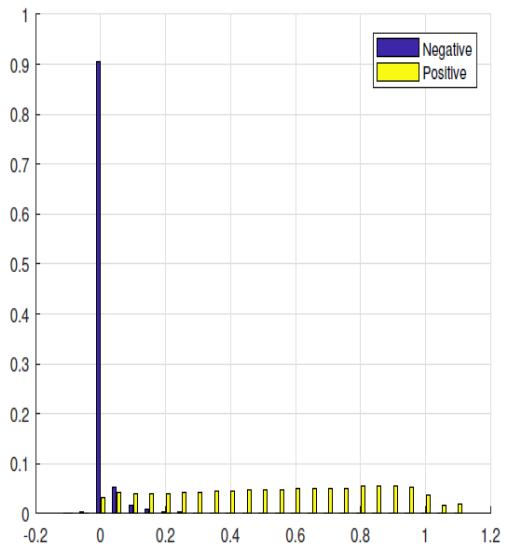
if $f(\mathbf{D}^{L'}) == f(\mathbf{D}^L)$ or $\mu \leq 2^{-20}$ then

Termination.

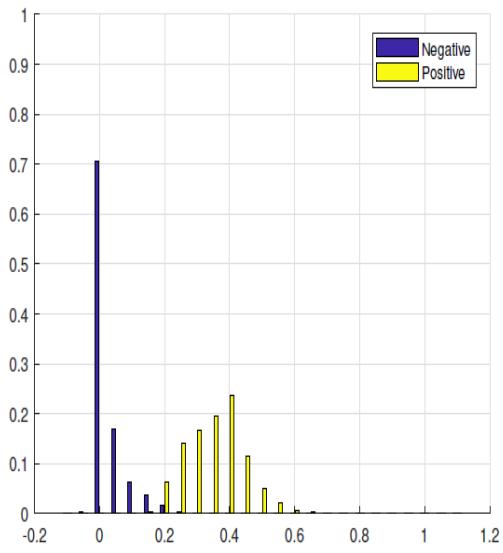
Update \mathbf{D}^L with $\mathbf{D}^{L'}$.



تنظیم دقیق نمایندگان معنایی برای امتیازهای حاشیه گسترش یافته



(الف)



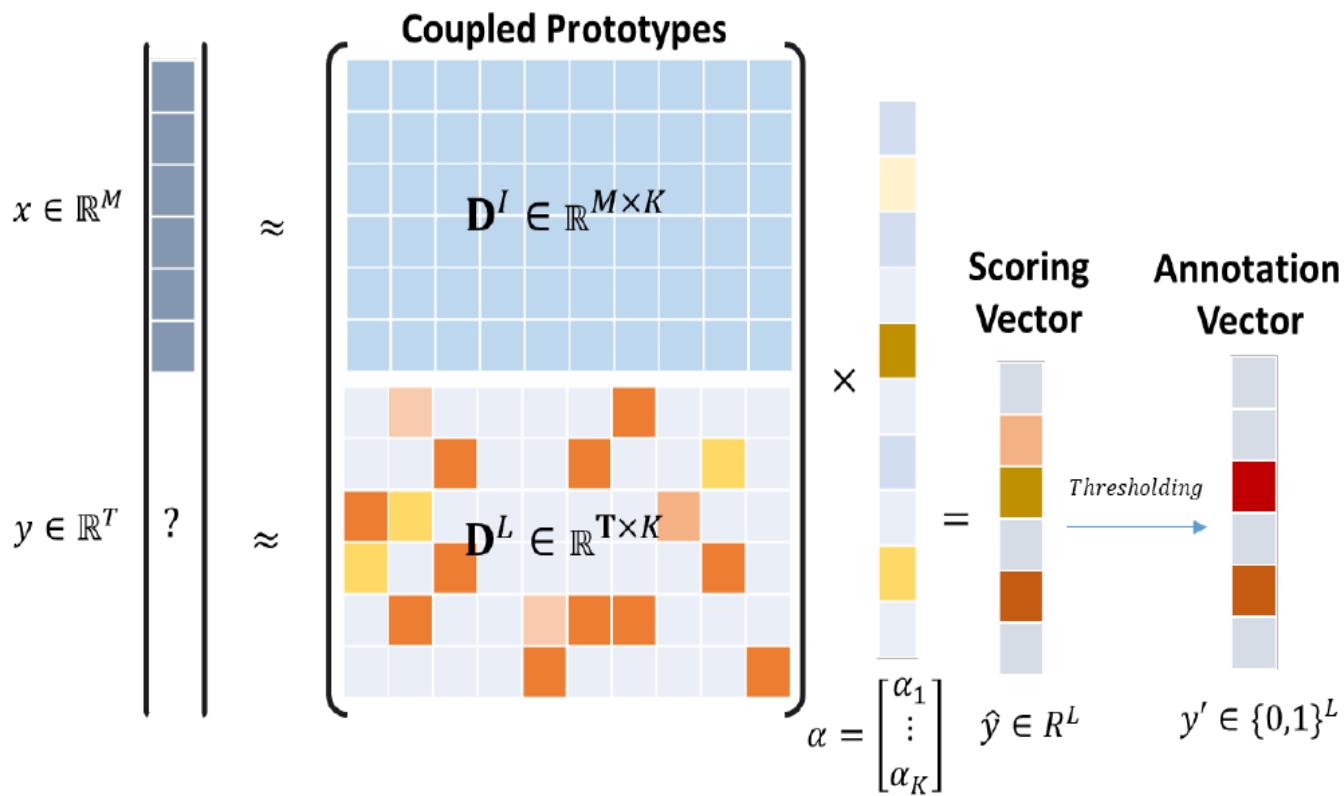
(ب)

شکل (۱۷) توزیع امتیازها برای برچسب‌های مثبت و منفی. (الف) تولید شده مبتنی بر مرحله آموزش نماینده توأم. (ب) بعد از به کارگیری روش حاشیه گسترش یافته

$\tau_{optimal \rightarrow F-measure Based}$

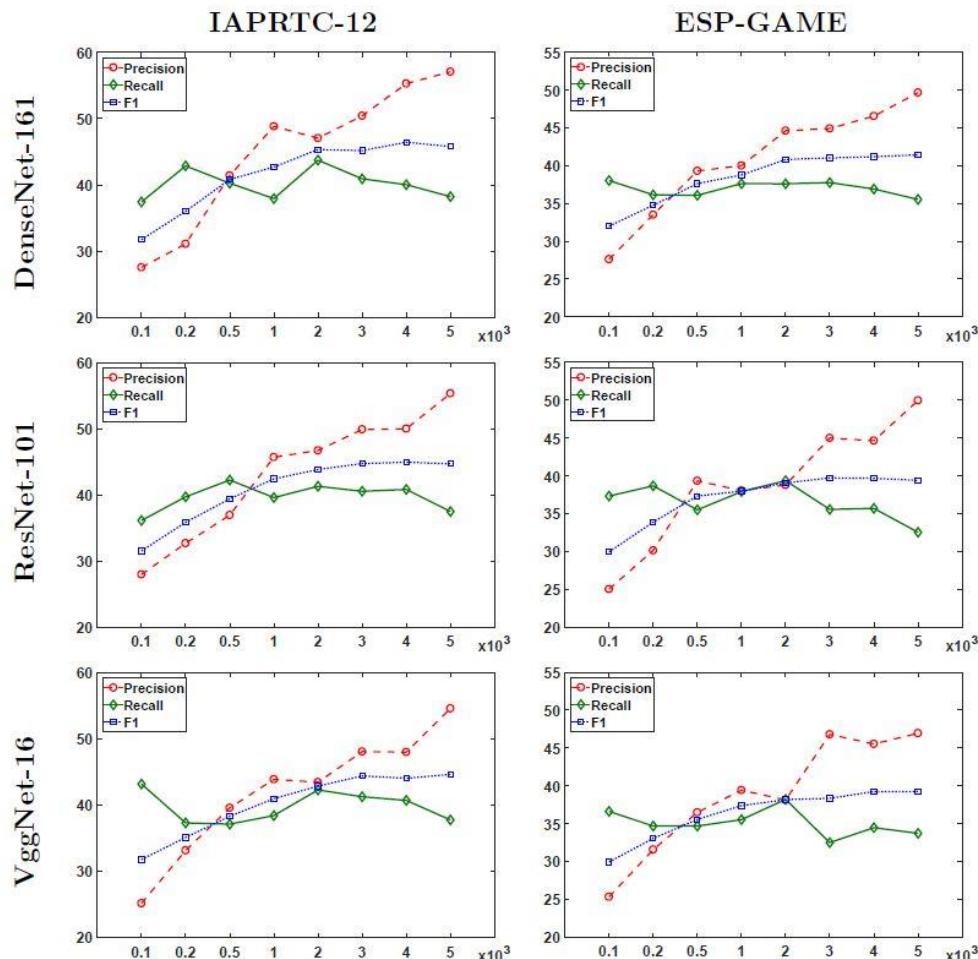


روش پیشنهادی برای نسبت دهی برچسب به تصویر آزمون ورودی



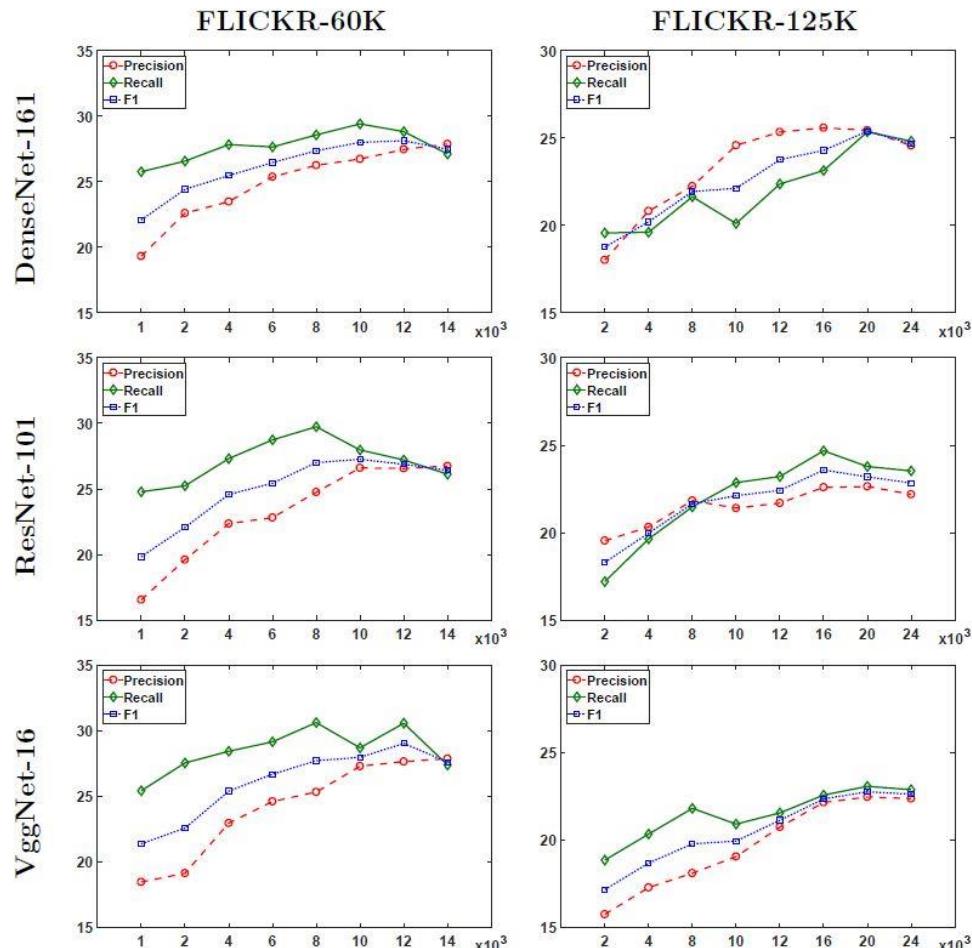


آنالیز حساسیت تعداد نمایندگان در کنار بررسی ویژگی های مختلف



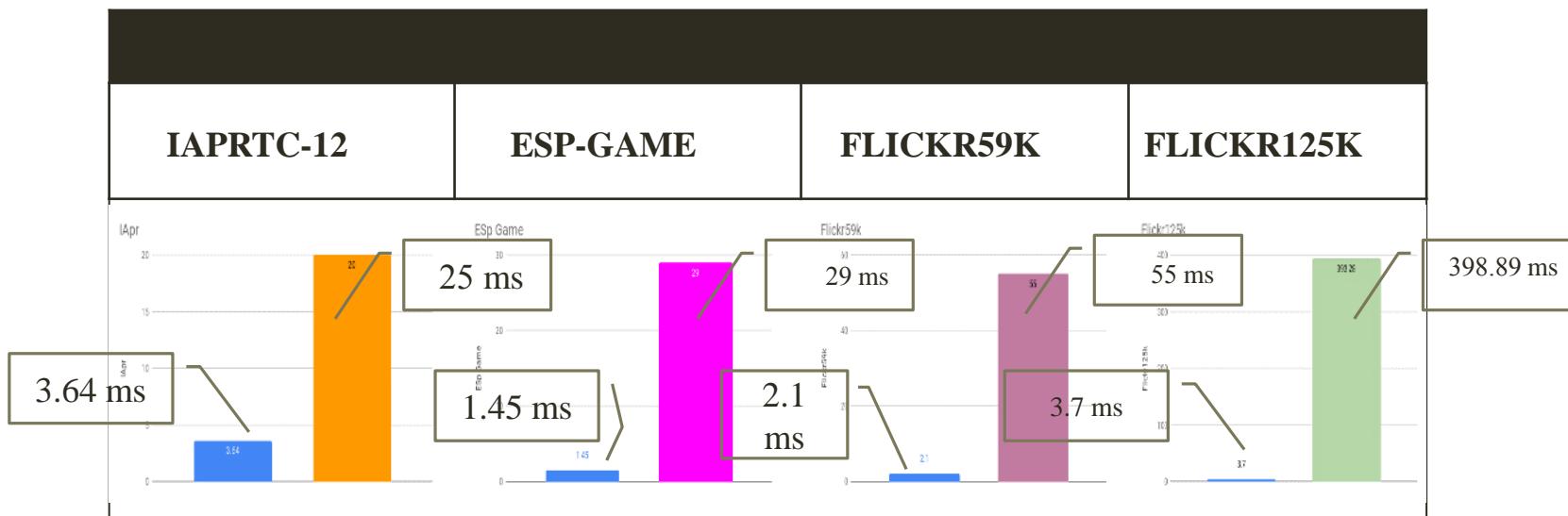


آنالیز حساسیت تعداد نمایندگان در کنار بررسی ویژگی های مختلف



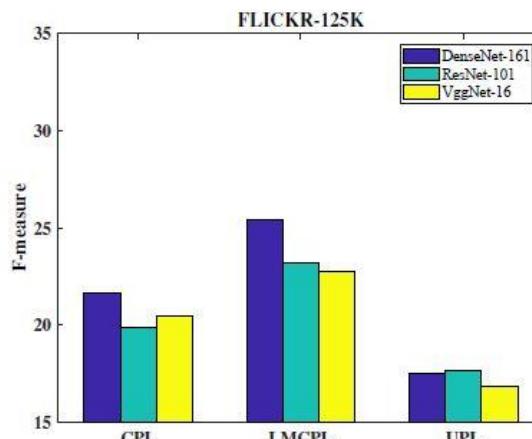
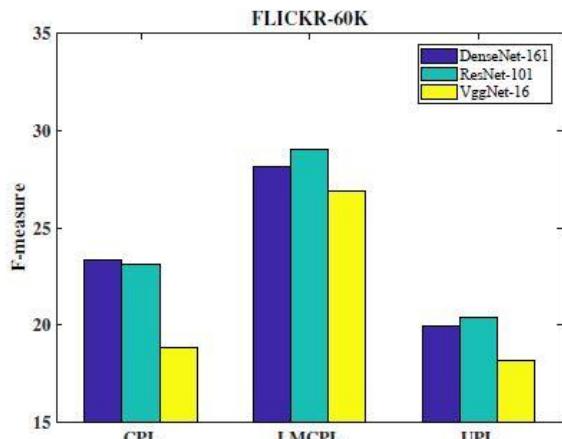
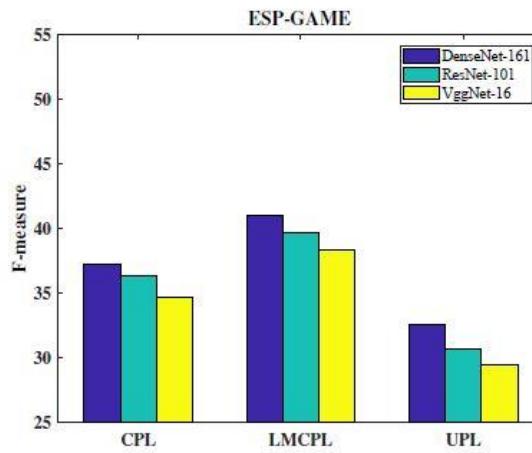
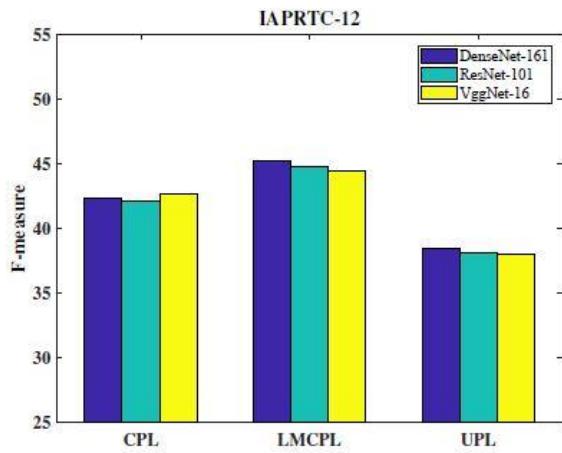


مقایسه سرعت برچسب زنی روش پیشنهادی با روش پایه





مقایسه آموزش LMSPL ، تک، تواناً





نتیجه گیری

- هدف از برچسب زنی پیش بینی برچسب هایی برای توصیف معنایی یک تصویر مبتنی بر محتوای بصری آن است.

← برچسب به کل تصویر تعلق می گیرد نه ناحیه خاصی

- روش های مبتنی بر جستجو مسئله برچسب زنی را مانند بازیابی بررسی کرده ازین جهت رویکرد ساده ای دارند، در عین حال کارایی بالاتری از بقیه روش ها دارند.

- به دلیل وجود شکاف معنایی نیازمند استخراج ویژگی های مناسب هستیم

← روش های مبتنی بر یادگیری ژرف

- عدم مقیاس پذیری سامانه های برچسب زنی

← استفاده از روش هایی که روند پیچیده ای در زمان تست ندارند.



نتیجه گیری

شباهت تصویر ورودی به پایگاه داده : مقایسه صرفاً تصاویر منتخب آموزشی با هر تصویر ورودی طی بررسی انجام شده، تعداد تصاویر آموزشی درگیر شده جهت مقایسه و پیدا کردن شباهت با تصویر ورودی بسیار کم است

- دو رویکرد پیشنهادی :
- 1. کاهش فضای جستجو
- 2. خلاصه سازی پایگاه داده به نمایندگان
- خلاصه سازی با استفاده از انتشار برچسب ها
- آموزش نماینده معنایی حاشیه گسترش یافته

مزیت اساسی این روش، کنار گذاشتن کل پایگاه داده آموزشی و استفاده از پایگاه داده جایگزین و کوچکتر معايب، شناس بیشتر برای برچسب های با فراوانی بالا طی انتشار برچسب

مبتنی بر بازنمایش تنک آموزش هم زمان تعداد محدود از نمایندگان بصري و برچسب های معنایي
مرتبه شن بازنمایش تنک آموزش هم زمان تعداد محدود از نمایندگان بصري و برچسب های معنایي
دوش برچسب زنی مبتنی بر بازسازی و نسبت دادن برچسبها به تصویر ورودی
از مون چالش، بيش بر از اش نماینده های معنایي بر روی داده ای آموزشی
چالش، استفاده از رویکرد حاشیه گسترش یافته جهت تعمیم پذیری بهتر



پیشنهادها

- ترکیب یا بازطراحی سامانه‌های پیشنهادی در رساله سبب کارائی بالاتر برچسبزنی برای شرکت داده شدن هر چه بیشتر برچسب‌های کمیاب با شанс برابر با برچسب‌های فراوان با مقیاس‌پذیری بسیار بهتری خواهد شد.
- برچسب‌های با همپوشانی بالا، ارزیابی مناسب‌تر ← محاسبه گستردگی میان برچسب‌های پیش‌بینی شده و میزان مفید بودن برچسب برای تصویر، مثلاً استفاده از بهره اطلاعاتی برچسب در میان برچسب‌های پیش‌بینی شده (Information Gain)
- کارایی بالاتر حافظه مصرفی، در روش پیشنهادی خلاصه سازی پایگاه داده، به جای ساخت گراف از کل مجموعه آموزشی، تقسیم گراف به چندین مولفه، خلاصه سازی بر روی هر مولفه به طور مجزا
- Flickr، تغییر یا تصحیح مجموعه داده ساخته شده جهت بهبود کارایی سامانه برچسب زنی ناکارآمدی جستجو میان نمایندگان برای مجموعه دادگان بزرگ: استفاده از کدگذاری و درهم سازی به جای جستجوی کامل میان نمایندگان

با تشکر از توجه شما

Any Question?

