

اللَّهُمَّ احْمِ
الْحَمْرَيْنِ آمِينَ!

“

عنوان پایان نامه: بازیابی تصویر ریزدانه‌ای مبتنی بر محتوا

ارائه دهنده: سید نیما سید آقا یزدی
استاد راهنما: دکتر کامبیز رهبر

”

سر فصل مطالب

فصل ۵
نتیجه‌گیری و پیشنهادات
آینده

فصل ۴
نتایج آزمایشگاهی

فصل ۳
روش پیشنهادی

فصل ۲
مروری بر پیشینه پژوهش

فصل ۱
کلیات پژوهش



تعریف مساله

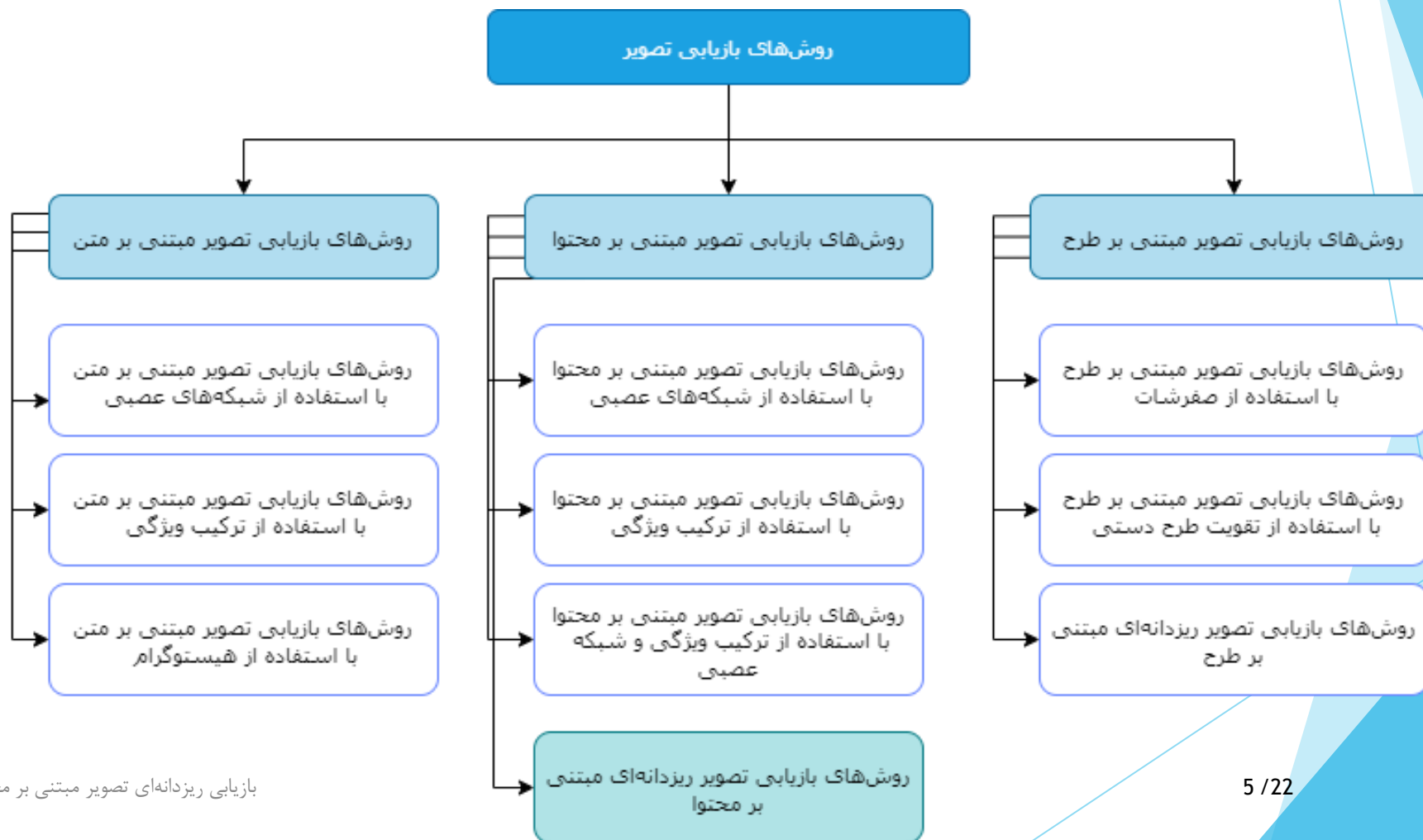
- تصاویر نقش مهمی در انتقال اطلاعات دارند. با پیشرفت سریع فناوری رایانه، میزان داده‌های تصاویر دیجیتال به سرعت در حال افزایش است. نیاز اجتناب‌ناپذیری به روش‌های کارآمدی وجود دارد که می‌تواند به جستجو و بازیابی اطلاعات بصری تصاویر کمک کند. بازیابی تصاویر از آن‌رو حائز اهمیت است که حجم زیادی از محتوای در دسترس را شامل می‌شود. با توجه به آنکه پردازش متن گاهی با کژتابی همراه است
- یکی از مهم‌ترین شاخه‌های پردازش تصویر، بازیابی تصاویر می‌باشد. بازیابی تصویر همان یافتن تصاویر مشابه از مجموعه بزرگی از داده‌های تصویری است. دو رویکرد اصلی برای حل مسئله بازیابی تصویر عبارت‌اند از: مبتنی بر متن و مبتنی بر محتوا. در رویکرد مبتنی بر محتوا.
- روش‌های حاضر دارای چند مشکل هستند و مورد اول اینکه در روش‌های مرسوم، تمرکز الگوریتم‌ها، روی مقایسه میزان تفاوت بخش‌های مختلف تصویر است. هنگامی که جستجوی ما داخل یک طبقه‌بندی ریزتر انجام گیرد، بازیابی ریزدانه‌ای نامیده می‌شود.

“

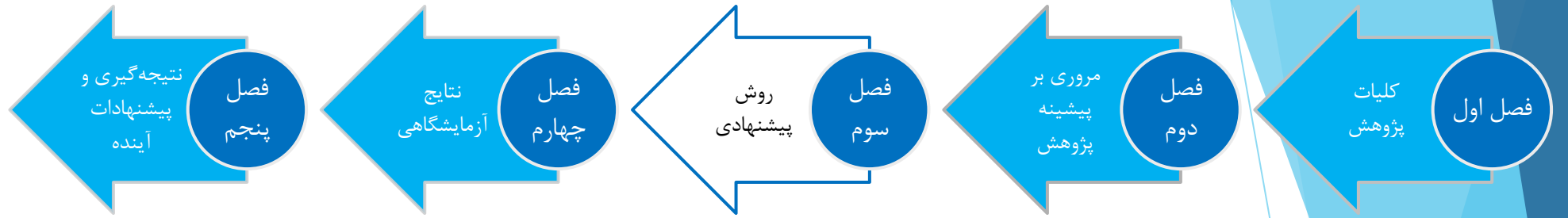


کاربردها

- پزشکی (عکس‌های CT-Scan)
- گیاه‌شناسی (طبقه‌بندی و تشخیص)
- جانورشناسی (تعیین نژاد جانوران)
- فروش (ارتباط بین انبار و پر کردن قفسه‌های سوپرمارکت)
- هنر (تشخیص سبک ادبی، زیبایی‌شناسی)



“



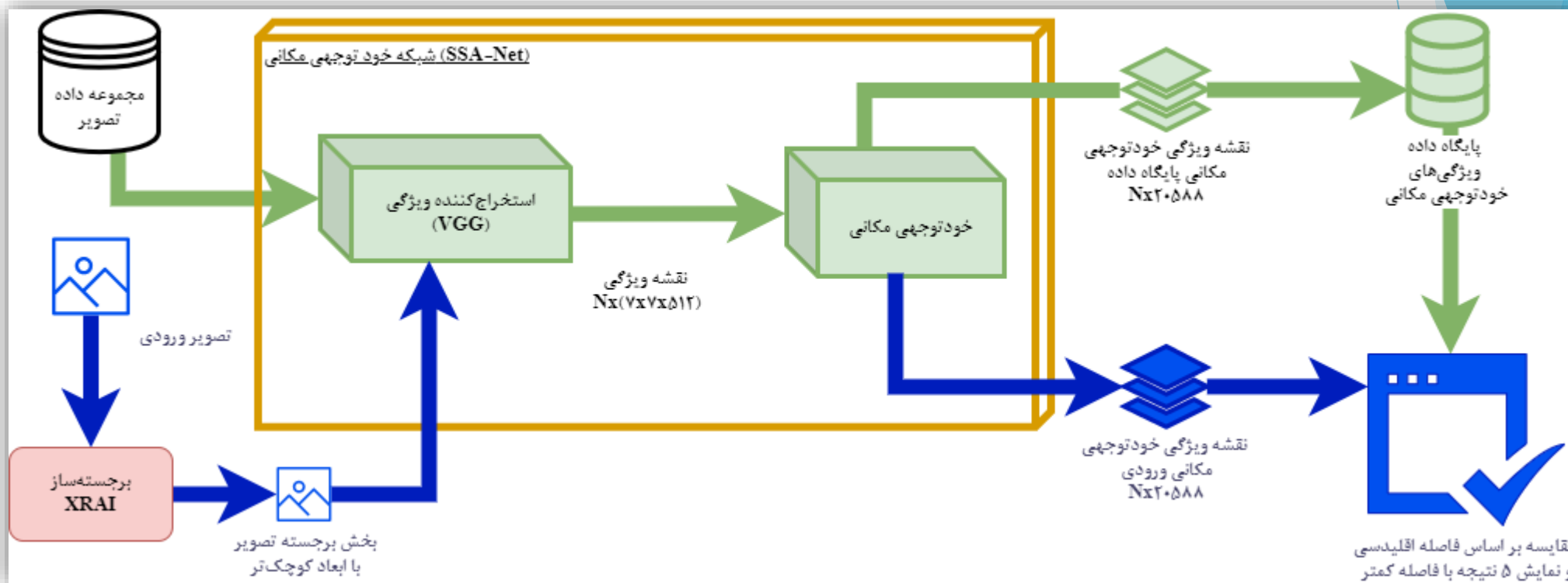
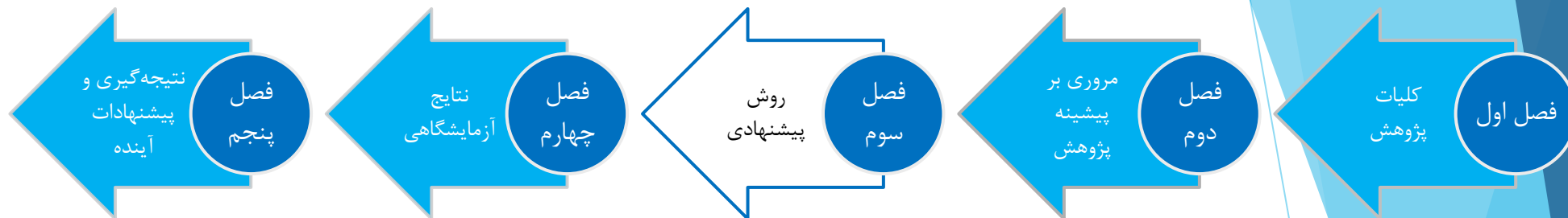
شبکه خودتوجهی مکانی از مکانیسم خودتوجهی برای دستیابی به ویژگی‌های تصویر استفاده می‌کند. نواحی مورد توجه بیشتر تقویت شده و امکان بررسی تفاوت‌ها را ساده‌تر می‌کند.

”

این شبکه از سه جزء اصلی تشکیل شده است:

- استخراج‌کننده ویژگی
- ماژول شبکه خودتوجهی مکانی
- بازیابی تصویر
- مکانیسم برجسته‌سازی

“

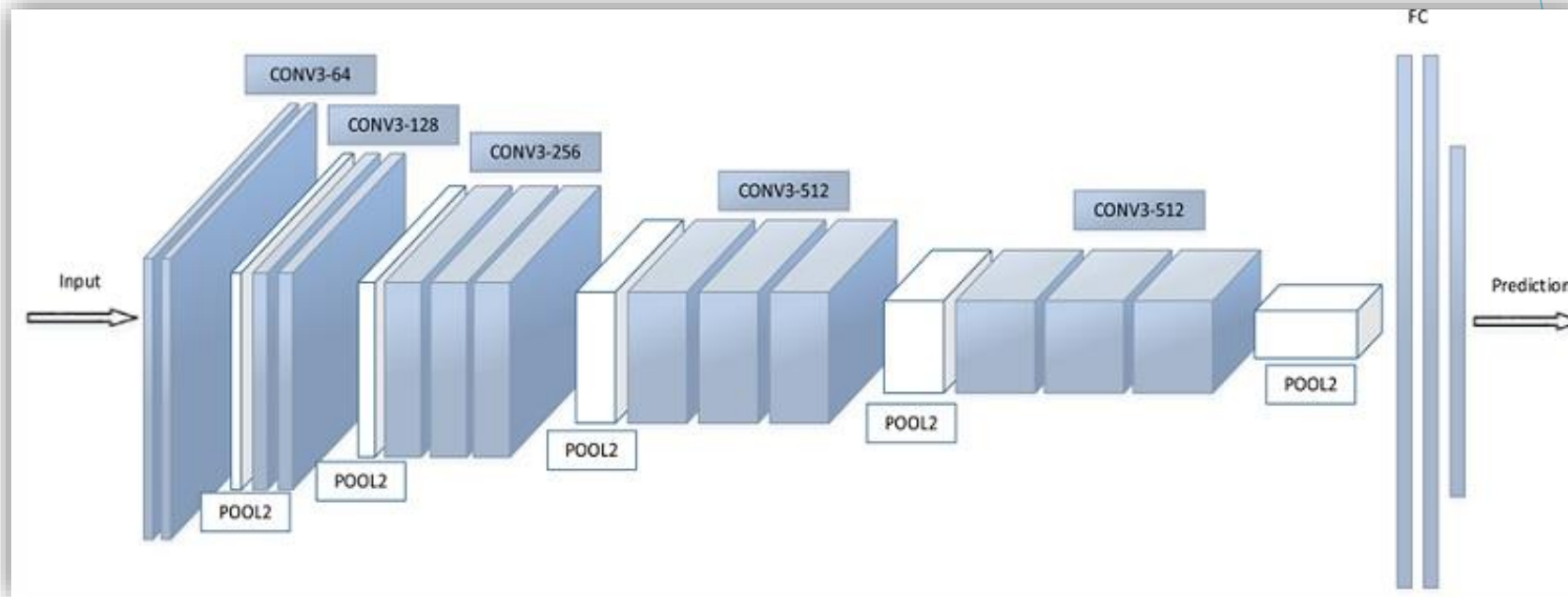


“



استخراج کننده ویژگی:

برای مقایسه منصفانه با سایر روش های پیشرفته، از VGG-16 از پیش آموزش دیده بر روی مجموعه داده ImageNet استفاده می شود.



“



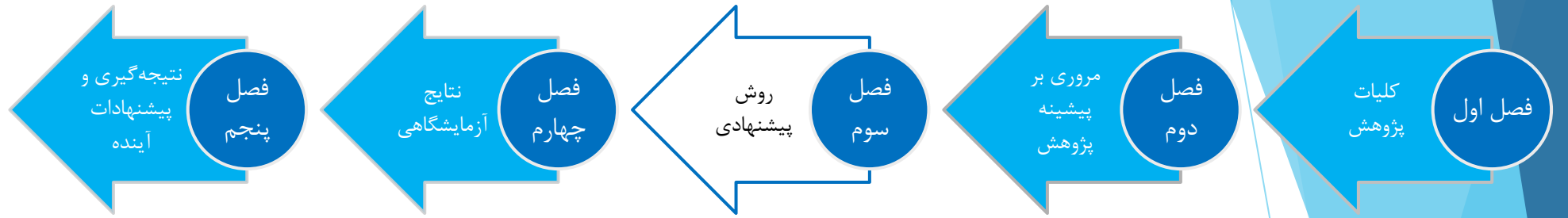
استخراج کننده ویژگی:

برای استخراج اولیه، سه لایه آخر که کاملاً متصل هستند حذف می شوند

ویژگی نقشه ها از تصاویر ورودی برای یک تصویر ورودی X یک مجموعه داده، خروجی نقشه ویژگی از لایه کانولوشنی نهایی گرفته می شود. این فرآیند به صورت نشان داده شده است

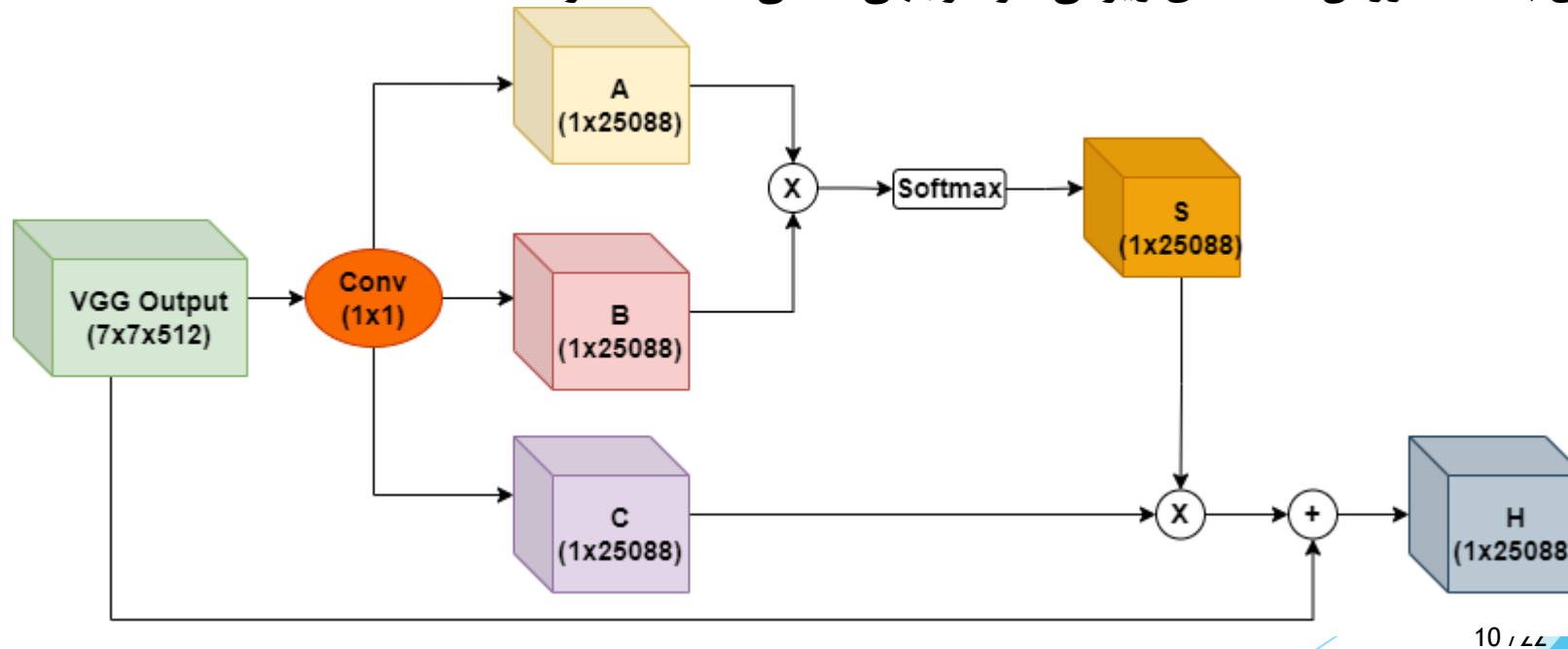
$$F = VGG(X)$$

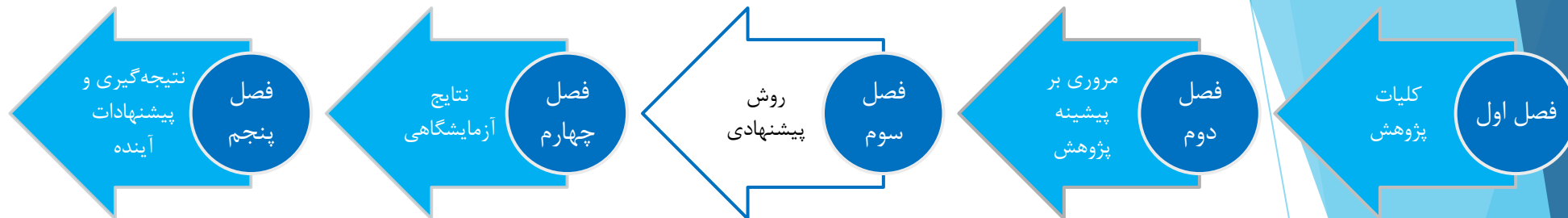
به طور خاص، استخراج کننده ویژگی یک تصویر ورودی X را به یک نقشه ویژگی ابعادی $F \in R^{H \times W \times K}$ نگاشت می کند، که در آن H ، W و K نشان دهنده ارتفاع مکانی، عرض مکانی، تعداد کانال ها/کرنل حاوی کانال هستند. این روند به ترتیب تا آخرین لایه پیش می رود.



ماژول خودتوجهی مکانی:

ماژول خودتوجهی مکانی از مکانیزم خودتوجهی پیشنهاد شده استفاده می‌کند که توجه محلی را از طریق یک تابع softmax جمع می‌کند. این ایده گسترش می‌یابد تا به موقعیت‌های پیکسل مکانی ویژگی‌های اصلی توجه شود و از تجمیع ویژگی‌ها برای به دست آوردن نقشه‌های ویژگی خودتوجهی مکانی استفاده شود.





ماژول خودتوجهی مکانی:

باتوجه به نقشه های ویژگی اولیه $F \in R^{H \times W \times K}$ به دست آمده از استخراج کننده ویژگی، ابتدا سه نقشه ویژگی جدید B ، A و C با استفاده از کانولوشن 1×1 تولید می شود.

$\{A, B, C\} \in R^{H \times W \times K}$ همان ابعاد فضای F را داراست. سپس A و B و C را به $R^{N \times K}$ تغییر شکل می یابد، که در آن $N = H \times W$ تعداد پیکسل ها است. سپس، ضرب عناصر بین A و ترانهاده B محاسبه می شود. "softmax" از نظر مکانی برای محاسبه نقشه خودتوجهی مکانی اعمال می شود $S \in R^{N \times N}$ که:

$$S_{ij} = \frac{\exp(A_i \otimes B_j)}{\sum_{i=1}^N \exp(A_i \otimes B_j)}$$

که در آن \otimes نشان دهنده ضرب عنصر است. S_{ij} نشان می دهد که چگونه شبکه تاثیر آمین موقعیت مکانی را بر موقعیت مکانی زمین اندازه گیری می کند.

“



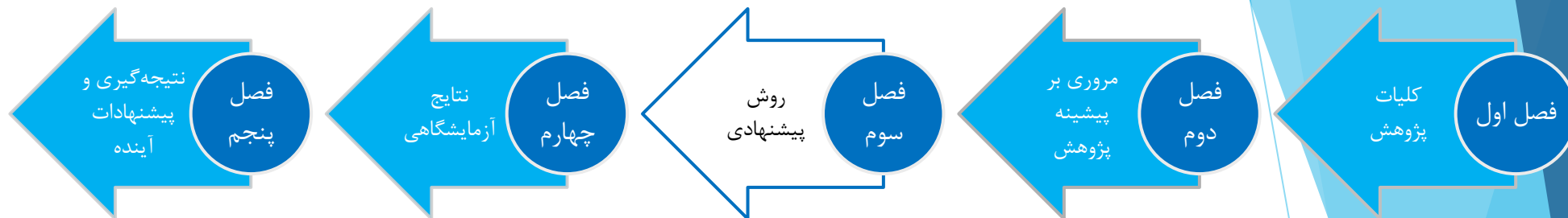
ماژول خودتوجهی مکانی:

در نهایت، یک مکانیسم تجمیع ویژگی برای بررسی تأثیر مناطق خودتوجهی مکانی در همه موقعیت‌ها در نقشه ویژگی اصلی از طریق معادلات پیاده‌سازی می‌شود:

$$H_j = \sum_{i=1}^N (s_{ij} C_i) \oplus F_j$$

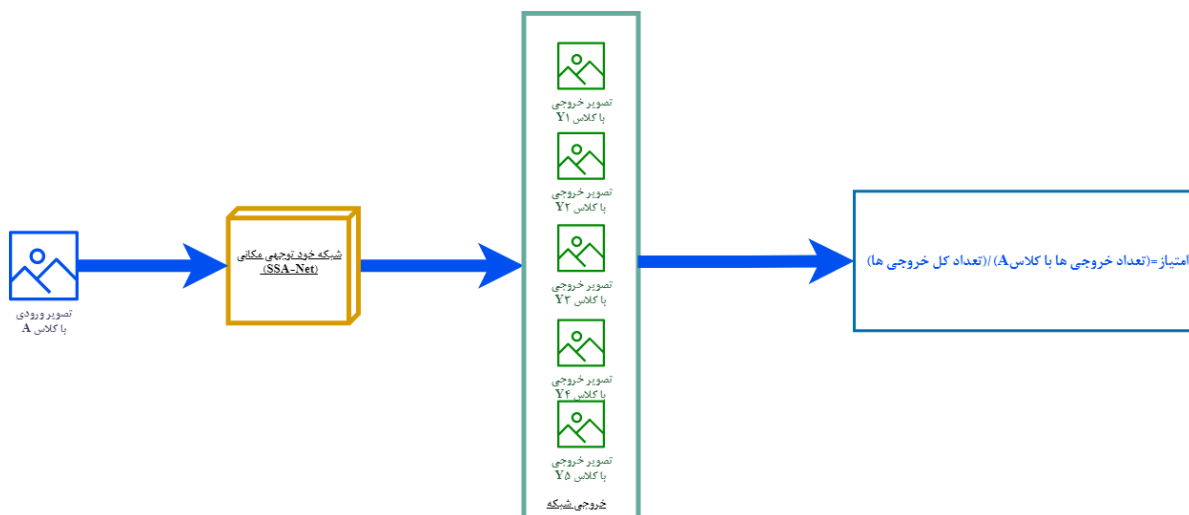
ویژگی‌های به دست آمده توسط H_j نشان‌دهنده یک تجمع کلی از نمای زمینه‌ای بر اساس نقشه‌های خودتوجهی مکانی است. مجموعه این ویژگی‌ها به عنوان یک پایگاه داده ذخیره می‌شوند.

“



بازیابی تصویر:

در این بخش یک تصویر به عنوان ورودی به شبکه داده می شود و نقشه ویژگی های خودتوجهی مکانی آن به دست می آید. سپس با نقشه های ویژگی ذخیره شده در شبکه خودتوجهی مکانی ، با استفاده از معادله زیر مقایسه می شوند.

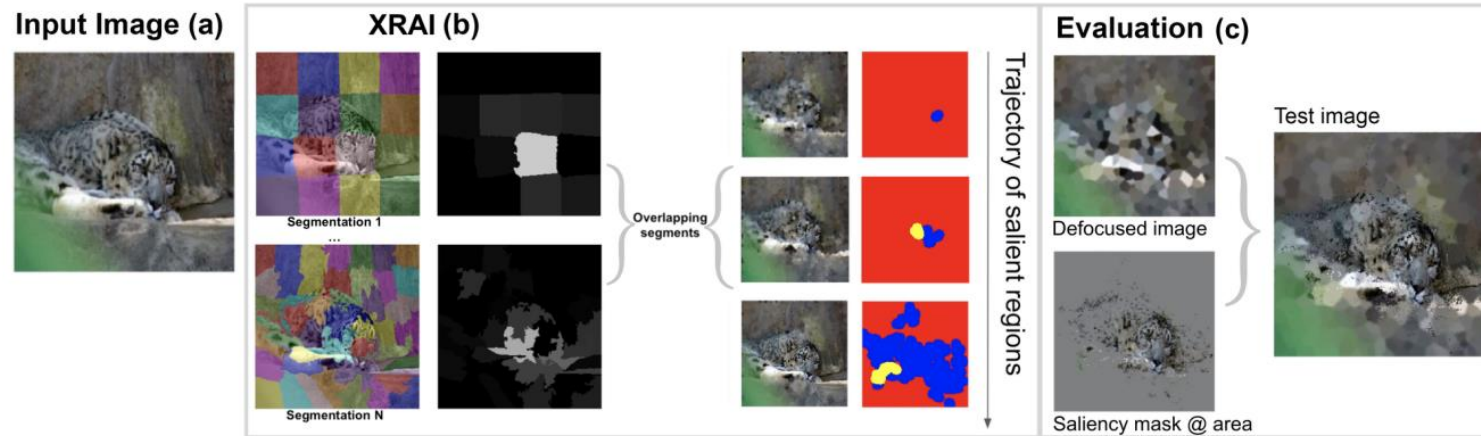


$$D(X, Y) = \sqrt{\sum_{i=1}^{N=20588} (X_i - Y_i)^2}$$

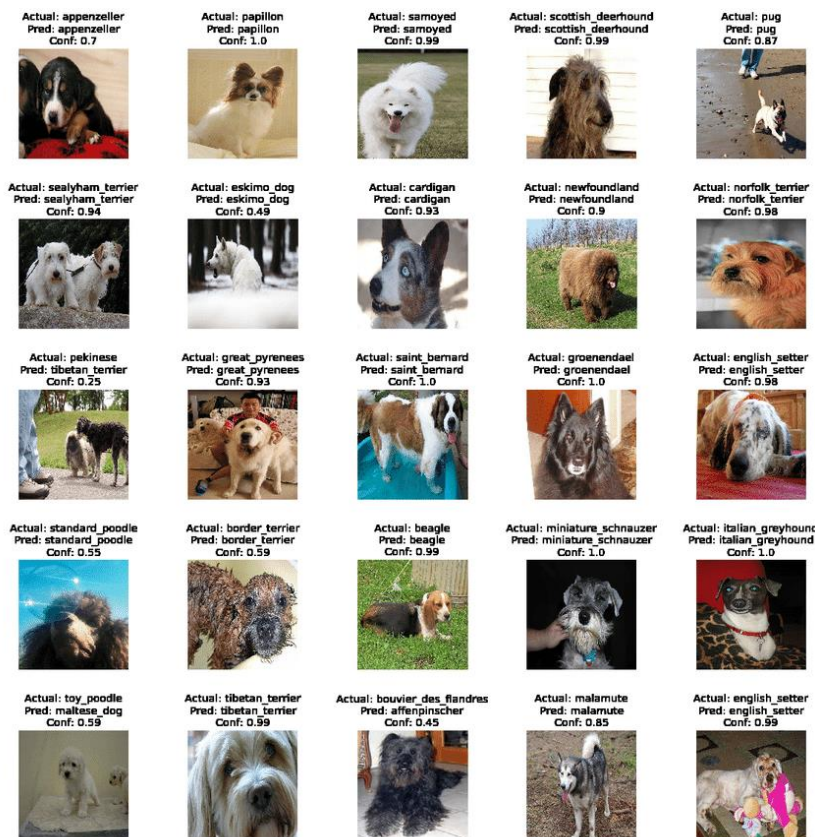


مکانیسم برجسته سازی (Saliency):

به عنوان راهکار پیشنهادی از روش XRAI برجسته سازی استفاده می گردد که مناطق پراهمیت تصویر را شناسایی کرده و موجب می شود یک قاب از تصویر ورودی را استخراج کنیم. بدین صورت بخش هایی از تصویر که دارای اهمیت کمتری هستند مورد پردازش قرار نمی گیرند و عملکرد کلی شبکه بهبود می یابد.



“



معرفی پایگاه‌های داده:

مجموعه داده Stanford Dogs شامل تصاویری از ۱۲۰ نژاد سگ از سراسر جهان است. این مجموعه داده با استفاده از تصاویر و حاشیه نویسی از ImageNet برای طبقه بندی تصاویر ریز دانه ساخته شده است. این پایگاه داده، در ابتدا برای دسته بندی تصاویر دانه ریز جمع آوری شد، یک مشکل چالش برانگیز وجود داشت، زیرا برخی از نژادهای سگ ویژگی های تقریباً یکسانی دارند یا از نظر رنگ و سن متفاوت هستند. در آزمایش پیش رو، این پایگاه داده از آن جهت که ساختار سلسله مراتبی داشته و تا حد بالایی رفتارهای ریزدانه ای ارائه می دهد، انتخاب شده است.

“



معرفی پایگاه‌های داده:

”
برای راستی‌آزمایی بهتر روش پیشنهادی، مجموعه‌داده CUB_200_2011 نیز مورد بررسی قرار گرفته‌است. مجموعه داده (CUB-200-2011) Caltech-UCSD Birds-200-2011 پرکاربردترین مجموعه داده برای کار دسته بندی بصری ریز دانه است. این شامل ۱۱۷۸۸ تصویر از ۲۰۰ زیرمجموعه متعلق به پرندگان، ۵۹۹۴ تصویر برای آموزش و ۵۷۹۴ تصویر برای آزمایش است.

“



معیارهای ارزیابی

معیار این آزمایش فاصله اقلیدسی بین نقشه ویژگی خودتوجهی مکانی تصویر ورودی و نقشه ویژگی خودتوجهی مکانی تصاویر پایگاه داده، می باشد.

”

$$D(X, Y) = \sqrt{\sum_{i=1}^{N=20588} (X_i - Y_i)^2}$$

که در آن X نقشه ویژگی خودتوجهی مکانی تصویر ورودی و Y نقشه ویژگی خودتوجهی مکانی هر تصویر از پایگاه داده است. سپس فاصله های به دست آمده، که هر کدام نگاشتی به تصویری از پایگاه داده دارند، به صورت نزولی مرتب شده و ۵ نتیجه برتر بازیابی می شود. خروجی سیستم بر اساس کلاسی که بیشترین احتمال را در بین این ۵ نتیجه دارد، تعیین می گردد. امتیاز هر بازیابی تصویر بر اساس فرمول زیر تهیه می گردد.

$$Score = \frac{Count\ Of\ True\ Labels}{Count\ Of\ All\ Results}$$

“



معیارهای ارزیابی

به عنوان مثال برای تصویر I با کلاس A که نتایج آن شامل $Results=[A,A,B,A,A]$ می باشند، داریم:

”

$$Score(I) = \frac{Count(Results\ with\ A)}{Count(Results)}$$

سپس عملکرد نهایی سیستم بر اساس فرمول زیر محاسبه خواهد شد:

$$Score = \frac{\sum_{i=1}^N Score(I_i)}{N}$$

که در آن N تعداد نمونه های تستی و I هر کدام از تصاویر تستی می باشد.

“



نتایج ارزیابی:

”

Input	Outputs					Score
						80%
Lhasa	Lhasa	Lhasa	Lhasa	Lhasa	valley terrier	
						80%
Irish setter	Irish setter	Irish setter	Irish setter	Irish setter	flat coated retriever	
						100%
Pomeranian	Pomeranian	Pomeranian	Pomeranian	Pomeranian	Pomeranian	
						60%
Afghan_hound	bloodhound	Afghan_hound	bloodhound	Afghan_hound	Afghan_hound	
						100%
otterhound	otterhound	otterhound	otterhound	otterhound	otterhound	
						100%
Alghan_hound	Alghan_hound	Alghan_hound	Alghan_hound	Alghan_hound	Alghan_hound	
						60%
Pekinese	Pekinese	Shih-Tzu	Shih-Tzu	Pekinese	Pekinese	
						100%
Borneo_mountain	Borneo_mountain	Borneo_mountain	Borneo_mountain	Borneo_mountain	Borneo_mountain	
						100%
chow	chow	chow	chow	chow	chow	
						100%
clumber	clumber	clumber	clumber	clumber	clumber	

Input	Outputs					Score
						56%
Wren	Warbler	Wren	Hummingbird	Wren	Wren	
						56%
Woodpecker	Woodpecker	Woodpecker	Woodpecker	Flicker	Woodpecker	
						56%
Sparrow	Thrush	Sparrow	Sparrow	Sparrow	Sparrow	
						56%
Vireo	Vireo	Vireo	Warbler	Vireo	Vireo	
						56%
Gull	Gull	Grebe	Gull	Gull	Gull	
						56%
Grebe	Grebe	Grebe	Cormorant	Grebe	Grebe	
						56%
Blackbird	Blackbird	Blackbird	Blackbird	Blackbird	Crow	
						56%
Gull	Gull	Gull	Gull	Gull	Gull	
						56%
Gull	Gull	Gull	Gull	Gull	Gull	
						56%
Grebe	Grebe	Grebe	Grebe	Grebe	Grebe	

“



مقایسه نتایج ارزیابی:

Method for Stanford Dogs	Score
FCAN	84.5%
PDFR	71.9%
PC-DenseNet-161	83.6%
HDWE	79.6%
EfficientNet-B0	61.2%
PC	61.9%
SSA	86%
SSA with XRAI Saliency	88%

Method for CUB_200_2011	Score
PDFR	%82.6
HDWE	%84.3
SSA	%85
SSA with XRAI Saliency	%87



- یکی از مهم‌ترین شاخه‌های پردازش تصویر، بازیابی تصاویر می‌باشد. این شاخه از علم پردازش تصویر، برای اولین بار با رویکرد مبتنی بر متن معرفی گردید. پس از آن رویکردی متفاوت با عنوان مبتنی بر محتوا معرفی گردید و در حوزه‌های پزشکی، گیاه‌شناسی، جانورشناسی، فروش و هنر مورد استفاده قرار گرفت.
- بازیابی تصویر در سه رویکرد مبتنی بر متن، مبتنی بر محتوا و مبتنی بر طرح، دنبال می‌گردد. هر کدام از این رویکردها در زیرروش‌های مختلفی دسته‌بندی می‌شوند که در نوع استخراج ویژگی، نوع پردازش ویژگی‌ها و طبقه‌بندی خروجی با هم تفاوت دارند.
- یکی از زیرشاخه‌های مبتنی بر محتوا، بازیابی ریزدانه‌ای تصویر است که تا یک دسته پایین‌تر تشخیص را جلوتر می‌برد و هدف آن کم کردن میزان اختلاف میان ویژگی‌های هر دسته از تصاویر می‌باشد. در بازیابی ریزدانه‌ای از بردار ویژگی، شبکه‌های عصبی مصنوعی و یا ترکیب این دو استفاده می‌شود.
- شبکه خودتوجهی مکانی بر اساس سه جزء اصلی ساخته شده است. یک ماژول جهت استخراج ویژگی، یک ماژول جهت خودتوجهی مکانی و یک ماژول جهت بازیابی تصویر و در نهایت یک ماژول به عنوان برجسته‌سازی تصویر ورودی. ماژول اول اطلاعات اصلی مورد نیاز را استخراج کرده و به ماژول دوم می‌دهد. در ماژول دوم با محاسبات روی بردار کانال‌های شده تصویر، میزان توجه به دست آمده و در ماژول آخر با سنجش فاصله ویژگی‌های تصویر ورودی با ویژگی‌های ذخیره شده خروجی را تعیین می‌کنیم. جهت حذف نواحی کم‌اهمیت تصویر از یک ماژول برجسته‌سازی نیز استفاده می‌شود.
- به عنوان تحقیقات آینده، پیشنهاد می‌شود رویکرد بازیابی تصویر ریزدانه‌ای مبتنی بر محتوا، با اضافه شدن ماژول یادگیری تقویتی ادامه یابد. می‌توان از یادگیری بدون نظارت استفاده کرد و داده‌هایی که ساختار سلسله مراتبی ندارند نیز بررسی گردند. سپس نتایجی که سیستم به عنوان خروجی برگردانده بررسی شده و امتیاز بگیرند. ممکن است در مواردی منجر به ساختن خوشه‌های جدید تصاویر شود و مجموعه تصاویر بتوانند با امتیازهای متفاوت دسته‌بندی سلسله مراتبی را ایجاد کنند.

با تشکر از توجه شما