



دانشگاه صنعتی شریف  
دانشکده مهندسی کامپیوتر  
سمینار کارشناسی ارشد گرایش هوش مصنوعی

عنوان:  
دسته‌بندی ریزدانه‌ای تصاویر  
Fine-grained Image Classification

نگارش:  
یاسر سوری  
۹۲۲۰۴۷۴۴

استاد راهنما:  
دکتر شهره کسایی

استاد ممتحن داخلی:  
دکتر محمد تقی منظوری شلمانی

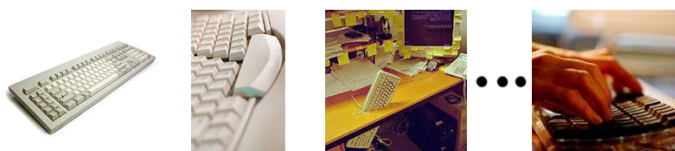
شهریور ۹۳

**چکیده:** دسته‌بندی تصاویر ریزدانه‌ای عبارت است از دسته‌بندی تصاویر در حالتی که دسته‌های مورد نظر همگی زیر دسته‌ی یک دسته‌ی کلی‌تر هستند. برای مثال برای زیر دسته‌ی کلی پرندگان ما می‌توانیم گونه‌های مختلف پرندگان را در نظر بگیریم. در این حالت خاص مسئله دسته‌ها معمولاً از نظر ظاهری بسیار به یکدیگر شبیه هستند به گونه‌ای که افراد غیر متخصص نمی‌توانند دسته‌ها را از یکدیگر تمایز دهند. در چنین شرایطی راه حل‌های ارائه شده برای مسئله دسته‌بندی تصاویر معمولی اکثراً نتایج خوبی کسب نمی‌کنند. لذا ارائه روش‌هایی جدید برای حل این مسئله لازم است. در این گزارش ابتدا به مرور روش‌های مهم در دسته‌بندی تصاویر معمولی و سپس به مرور روش‌های ارائه شده برای دسته‌بندی تصاویر ریزدانه‌ای می‌پردازیم. سپس روش انتخاب شده و دلایل انتخاب آن را بررسی می‌کنیم.

**واژه‌های کلیدی:** بینایی کامپیوتری، بازشناسی شیء، دسته‌بندی تصاویر، بازشناسی ریزدانه‌ای، دسته‌بندی تصاویر ریزدانه‌ای.

## ۱ مقدمه

در دسته‌بندی تصویر<sup>۱</sup> هر تصویر با توجه به محتوایش دسته‌بندی می‌شود. برای مثال آیا تصویر شامل شیء خودرو هست یا خیر. معمولاً در بینایی کامپیوتری مسئله بدین صورت است که تعدادی دسته مشخص را در نظر می‌گیریم (مثلاً انسان، خودرو، ساختمان، تلویزیون، صندلی، اسب و ...) سپس طبق چارچوب معمول یادگیری ماشین، توسط تعدادی تصویر شامل یکی از دسته‌ها (نمونه‌های مثبت) و تعدادی تصویر بدون شیء از آن دسته (نمونه‌های منفی) یادگیری برای آن دسته انجام می‌شود. در نهایت پس از یادگیری تمام دسته‌ها در مواجهه با تصویر جدید لازم است تشخیص دهیم که آیا شیء از هر کدام از آن دسته‌های مورد بررسی در تصویر وجود دارد یا خیر. برای نمونه به شکل ۱، ۱ توجه کنید. در این شکل داده‌های آموزشی و آزمایشی برای دسته‌بند دسته‌ی «صفحه کلید<sup>۲</sup>» از پایگاه داده Caltech۲۵۶ [۱] نشان داده شده است.



(آ) نمونه‌های مثبت، شامل شیء صفحه کلید



(ب) نمونه‌های منفی، بدون شیء صفحه کلید



(ج) تصویر جدید آزمایش

**شکل ۱، ۱:** نمونه‌ای از تصاویر آموزشی و آزمایشی سامانه دسته‌بند تصاویر برای دسته‌ی «صفحه کلید». انتخاب شده از پایگاه داده Caltech۲۵۶ [۱]. سامانه لازم است با مشاهده نمونه‌های مثبت ۱، ۱ و نمونه‌های منفی ۱، ۱ یادگیری را انجام داده و بتواند پاسخ دهد که در تصویر جدید آزمایشی ۱، ۱ آیا صفحه کلید وجود دارد یا خیر. برای این نمونه پاسخ مثبت است.



(ب) هواپیمای مسافربری



(آ) هواپیمای جنگنده

شکل ۲,۱: نمونه‌ای از دو دسته‌ی متفاوت ولی شبیه به هم از پایگاه داده Imagenet [۲]. مدل دسته‌بند علاوه بر توانایی مدل‌سازی تفاوت‌های داخل دسته‌ای، باید توانایی تمایز بین دسته‌های گاه شبیه به یکدیگر را داشته باشد.

چالش‌های اصلی این مسئله تنوع زیاد اشیاء درون هر کدام از دسته‌ها، نحوه‌ی عکس برداری و وجود اشیاء دیگر در تصویر است که باعث ایجاد تصاویری با تنوع بالا می‌شود. مدل کردن این تنوع مربوط به اشیاء هر دسته باید همزمان با توانایی تمایز بین دسته‌های مختلف باشد. برای نمونه شکل ۲,۱ تصویری از دو دسته مختلف مربوط به پایگاه داده Imagenet [۲] را نمایش می‌دهد. مدل دسته‌بند باید توانایی تمایز بین این دو دسته شبیه به هم را داشته باشد.

اگر در دسته‌بندی تصویر، دسته‌های مورد بررسی زیر دسته‌ی ۳ یک دسته‌ی کلی‌تر باشند (مانند گونه‌های مختلف پرندگان، مدل‌های مختلف خودروهای سواری و انواع مختلف هواپیماها)، آنگاه مسئله را «دسته‌بندی ریزدانه‌ای تصویر ۴» می‌نامند. در دسته‌بندی ریزدانه‌ای تصویر معمولاً شباهت دسته‌ها به یکدیگر بسیار زیاد است به نحوی که افراد غیر متخصص نمی‌توانند به راحتی این دسته‌ها را بازشناسی نمایند. برای نمونه در شکل ۳,۱ چند گونه‌ی مختلف از پرستوی دریایی<sup>۵</sup> متعلق به پایگاه داده CUB-۲۰۰-۲۰۱۱ [۳] نمایش داده شده است. همانگونه که می‌بینید با اینکه این نمونه‌ها به نحوی انتخاب شده‌اند که وضعیت مشابهی دارند، هنوز هم پیدا کردن ویژگی‌های تمایز دهنده بین گونه‌های مختلف کار بسیار سختی است و نیاز به تخصص دارد.



The elegant tern (د)



The common tern (ج)



The Artic tern (ب)



The Caspian tern (آ)

شکل ۳,۱: چهار گونه‌ی مختلف از پرستوهای دریایی متعلق به پایگاه داده CUB-۲۰۰-۲۰۱۱ [۳]. شباهت بسیار زیاد بین دسته‌های مختلف کار را حتی برای انسان‌های غیر متخصص بسیار سخت می‌کند.

روش‌های دسته‌بندی تصویر معمولی در مسایل دسته‌بندی ریزدانه‌ای اکثراً موفق نیستند (به بخش فولان مراجعه شود). دلیل اصلی این عدم موفقیت وجود ویژگی‌های بسیار اندک و شدیداً محلی تمایز دهنده<sup>۶</sup> برای دسته‌های ریزدانه ایست. برای مثال دو گونه‌ی elegant tern در شکل ۳,۱ و common tern در شکل ۳,۱ فقط در رنگ پا و شکل تاج با یکدیگر تفاوت دارند و در سایر اجزا غیر قابل تمایز هستند.

مسئله مطرح بوده است.

مسئله در حالت کلی بسیار مشکل است.

یک مسئله معکوس است.

برای مقایسه کارایی باشد.

به صورت یک مسئله تخمین حالت سیستم پویا<sup>۷</sup> در نظر گرفت.

$$p(X_t|I_t) \propto \underbrace{p(I_t|X_t)}_{\text{درستمایی}} \int \underbrace{p(X_t|X_{t-1})}_{\text{مدل جابجایی}} p(X_{t-1}|I_{t-1}) dX_{t-1} \quad (۱)$$

## References

- [1] G. Griffin, A. Holub, and P. Perona, “Caltech-256 object category dataset,” California Institute of Technology, Tech. Rep., 2007.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [3] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, “The Caltech-UCSD Birds-200-2011 Dataset,” California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

## واژه‌نامه

Discriminative<sup>۶</sup>  
Dynamic System<sup>۷</sup>

Subclass<sup>۳</sup>  
Fine-grained image classification<sup>۴</sup>  
Term<sup>۵</sup>

Image classification<sup>۱</sup>  
Computer - Keyboard<sup>۲</sup>