

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران) دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پروژه کارشناسی گرایش نرم افزار

پرسش و پاسخ بصری با استفاده از شبکه های عصبی کانولوشنی عمیق و بازگشتی

نگارش

على غلامي

استاد راهنما

دكتر محمد رحمتي

فروردین ۱۳۹۶



صفحه فرم ارزیابی و تصویب پایان نامه- فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تایید و تصویب پایان نامه موسوم به فرم کمیته دفاع- موجود در پرونده آموزشی- را قرار دهید.

نكات مهم:

- نگارش پایان نامه/رساله باید به زبان فارسی و بر اساس آخرین نسخه دستورالعمل و راهنمای تدوین پایان نامه های دانشگاه صنعتی امیرکبیر باشد.(دستورالعمل و راهنمای حاضر)
- رنگ جلد پایان نامه/رساله چاپی کارشناسی، کارشناسی ارشد و دکترا باید به ترتیب مشکی، طوسی و سفید رنگ باشد.
 - چاپ و صحافی پایان نامه/رساله بصورت پشت و رو(دورو) بلامانع است و انجام آن توصیه می شود.

به نام خدا

تاریخ: فروردین ۱۳۹۶

تعهدنامه اصالت اثر



اینجانب علی غلامی متعهد می شوم که مطالب مندرج در این پایان نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیر کبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نگردیده است.

در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایاننامه متعلق به دانشگاه صنعتی امیرکبیر میباشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخهبرداری، ترجمه و اقتباس از این پایان نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

على غلامى امضا

کسی که من اینگونه به اعتاد کلید قلیم را در دستانت می کذارم نان شادی ام را با تو قسمت می کنم • ان شادی ام را با تو قسمت می کنم به کنارت می نشینم وسربرشانهی تو اليخين آرام به خواب می روم ب



از پدرم که چون کوهی استوار و مادرم که چون دریای محبت در فراز و نشیب زندگی دلسوزانه همراهم بوده اند؛ از استاد بزرگوار جناب آقای دکتر محمد رحمتی که در کمال سعه صدر، با حسن خلق و فروتنی، رهنمون من شده اند؛

و از سایر عزیزانی که در کنارشان این نتیحه حاصل آمد؛

کمال تشکر و قدردانی را دارم.

علی غلامی فروردین ۱۳۹۶

چکیده

در این پروژه، مدلی مبتنی بر شبکه های عصبی کانولوشنی و بازگشتی عمیق، به منظور پرسش و پاسخ بصری ارائه شده است. در ک ماشین از تصاویری که به آن ارائه می شود، هیچ گاه با در ک انسان از تصاویر قابل قیاس نبوده است. به همین دلیل، پژوهش های انجام گرفته در بینایی ماشین، همواره در صدد بهبود این قابلیت در ماشین ها بوده است. برای این منظور، ویژگی های مهم تصاویر استخراج شده و سپس در مورد وجود یا عدم وجود اشیا مورد نظر در تصاویر تصمیم گیری انجام می گیرد. فراتر از این موضوع، قابلیت در ک ماشین از سوالات انسان درباره این تصاویر و پاسخ به این سوالات بر اساس ویژگی ها و دانش مستخرج از تصاویر است که به تازگی مورد توجه پژوهشگران عرصه ی بینایی ماشین قرار گرفته است.

واژههای کلیدی:

پرسش و پاسخ بصری، شبکه های کانولوشنی عمیق، شبکه های بازگشتی عمیق، پردازش زبان طبیعی، پردازش تصویر)

فهرست مطالب	عنوان
	۱ مق
ﻪﻟﻪﻟﻪ ١٠ مقدمه	
۱۰ مقدمه	
۳۰ مروری بر روند پژوهش های پیشین	
وری بر مطالعات گذشته	۲ مر
۱ مقدمه	-٢
۱-۱-۲ مغز چگونه تصاویر را درک می کند؟	
ش ارائه شده در پژوهش	۳ رو
۱۰ مقدمه	۳-
۲۰ بیان مشکلات موجود ۲۰ ۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰	-٣
مخصات یک پایان نامه و گزارش علمی	ئه ۴
۱۰ برخورداری از غنای علمی	-۴
۲۰ ارجاع بهموقع و صحیح به منابع دیگر	-۴
۳۰ سادهنویسی	-۴
۴ وحدت موضوع	-۴
۵ اختصار	-۴
۶۰ رعایت نکات دستوری و نشانه گذاری	
۷۰ توجه به معلومات ذهنی مخاطب	-۴
۸ رعایت مراحل اصولی نگارش	-4
معبندی و نتیجهگیری و پیشنهادات	۵ جا
۱۰ پیشنهادات	-Δ
و مراجع	منابع
ت	پيوس
مهی فارسی به انگلیسی	واژەنا
مهی انگلیسی به فارسی	وادُونا

فهرست اشكال

فهرست اشكال

صفحه

شكل

فهرست جداول

فهرست جداول

جدول

فهرست نمادها

نماد مفهوم \mathbb{R}^n n فضای اقلیدسی با بعد n کرہ یکه n بعدی \mathbb{S}^n M جمینهm-بعدی M^m M وی هموار روی M $\mathfrak{X}(M)$ (M,g) مجموعه میدانهای برداری هموار یکه روی $\mathfrak{X}^{\prime}(M)$ M مجموعه p-فرمیهای روی خمینه $\Omega^p(M)$ اپراتور ریچی Qتانسور انحنای ریمان \mathcal{R} تانسور ریچی ricمشتق لي L۲-فرم اساسی خمینه تماسی Φ التصاق لوى-چويتاي ∇ لاپلاسين ناهموار Δ عملگر خودالحاق صوری القا شده از التصاق لوی-چویتای ∇^* متر ساساكى g_s التصاق لوی-چویتای وابسته به متر ساساکی ∇ عملگر لاپلاس-بلترامی روی p-فرمها Δ

٥

فصل اول مقدمه

1-1 مقدمه

درک انسان از محیط اطراف خود، از ابتدای کودکی شکل می گیرد. اگر از یک کودک، درباره جهانی که برای او قابل رویت است، سوالی پرسیده شود، بلافاصله درباره ی این جهان جملاتی توصیفی با دقت بسیار بالا ارائه می کند. انتقال این قدرت به ماشین ها از این جهات مختلفی حائز اهمیت می باشد. یکی از جهات اصلی آن، کاربرد های بی شماری است که می توان از این قابلیت در امور مختلفی استفاده کرد. دوربین های نظارتی، خودروهای خودران و ... نمونه هایی از این کاربرد ها هستند. در این فصل ابتدا موضوع پژوهش را به طور کامل بیان کرده و اهمیت ارائه راهکار مناسب در این مورد را بررسی می کنیم. سپس رویکرد های مختلف را برای حل این مساله بیان می کنیم.

1-1 توضيح مساله

درک تصاویر به همراه درک سوالاتی که مرتبط با آن تصاویر پرسیده می شود، می تواند یکی از اساسی ترین قابلیت های یک سامانه مدیریت هوشمند تصاویر، خودروی خودران و یا یک سیستم بازیابی تصویر هوشمند باشد. برای این منظور نیاز است که ابتدا صحنه به نمایش درآمده توسط ماشین هضم گردد. بدین معنی که بدون درک درست از آنچه در تصویر موجود است، ساخت چنین سامانه ای امکان پذیر نخواهد بود.

$\Upsilon-1$ مروری بر روند پژوهش های پیشین

روش هایی که جهت حل مسائل پرسش و پاسخ بصری ارائه شده اند از تنوع بالایی برخوردار هستند. اگرچه، می توان اغلب این روش ها را در راستای حل چهار چالش مهم زیر در نظر گرفت:

- چالش استخراج ویژگی از تصاویر
- ۲. چالش استخراج ویژگی از متن(سوالات پرسیده شده)
- ۳. چالش ترکیب بردار های ویژگی مستخرج از تصویر و متن
 - ۴. چالش تولید پاسخ متناسب با ویژگی ها ترکیب شده

با تحول عظیمی که در سال ۲۰۱۲ با ارائه مدل کانولوشنی عمیق جهت استخراج ویژگی ارائه شد و نیز افزایش قدرت پردازشی ماشین ها، توجه به سمت شبکه های کانولوشنی جهت استخراج ویژگی از تصاویر بیشتر شد. با پیشرفت این مدل ها و ترکیب آنها با مدل های پردازش زبان طبیعی، تولید شرح بر تصاویر نیز مورد توجه قرار گرفت. از سال ۲۰۱۵ تا به اکنون، یکی از جالب ترین موضوعاتی که توجه پژوهشگران را به سمت خود جلب کرده است، موضوع پرسش و پاسخ بصری می باشد. این موضوع ارتباط تنگاتنگی با موضوع تولید شرح بر تصاویر دارد. به همین منظور، بسیاری از تکنییک های رایج در بحث شرح بر تصاویر، در موضوع پرسش و پاسخ بصری نیز مورد توجه قرار گرفته است.

فصل دوم مروری بر مطالعات گذشته

۱-۲ مقدمه

دراین بخش به بررسی راهکار های مختلف ارائه شده در موضوغ پرسش و پاسخ بصری می پردازیم. پرسش و پاسخ بصری، یکی از چالش های بزرگ روز در عرصه ی هوش مصنوعی می باشد. بسیاری از ایده های مطرح در حل این چالش ها، از نجوه ی عملکرد ذهن انسان مدل برداری شده است.

مغز چگونه تصاویر را درک می کند؟

تصاویر در مغز انسان به طرز خارق العاده ی پردازش می شوند. به گونه ای که در اولین نگاه می توان بیشترین اطلاعات موجود در تصویر را استخراج کرد و آنها را توصیف نمود. بعلاوه، قدرت مغز در تخلیل شنیدار و پاسخ به این تحلیل و نیز صحنه ی دریافتی از طریق بینایی، همواره قابل تامل و مطالعه می باشد.

این ایده که مغز قادر است تا حجم زیادی از اطلاعات را به سرعت پردازش کرده و در مورد آنها تصمیم اتخاذ کند، از جانب پژوهشگران مورد بررسی قرار گرفته است. به عنوان مثال [۲] پژوهشی است که در آن تعدادی از تصاویر به صورت دنباله ای به افرادی نشان داده می شود و نیز توصیفاتی از طرف آنها ارائه می گردد. پژوهشگران با انجام این آزمایش یی بردند که مغز قادر است در کمتر از ۲۰۰ میلی ثانیه به صحنه های دریافتی یاسخ دهد. در پژوهش [۱] آزمایش دیگری انجام شده است که از اهمیت بسیاری بر خوردار است. در پژوهش های قبلی، افرادی که تصاویر را توصیف می کردند، درباره موضوغ کلی تصاویر اطلاعاتی داشتند.اما در این آزمایش، تصاویر مختلفی از دنیای واقعی که محدود به شرایط خاضی نبوده اند، بدون ارائه پیش فرض درباره ی موضوع، به افراد نمایش داده شده و از آنها خواسته شده که تصویر را به بهترین شکل توصیف کنند. نتایج بدست آمده به صورت زیر می باشد: ۱. حداکثر زمان لازم برای مغز انسان به منظور درک صحنه، برابر با ۵۰۰ میلی ثانیه می باشد.

۲. این مدت زمان، برای صحنه های ساده و بدون پیچیدگی، به حدود ۱۰۰ میلی ثانیه می رسد.

فصل سوم روش ارائه شده در پژوهش

۱-۳ مقدمه

در این فصل باید روشی جهت پیاده سازی یا نحوه ی انجام تحقیقات صورت گرفته ارائه گردد. این روش باید جزییات ویژه ای را پوشش دهد، چالش های مطرح را حل و فصل و نیز نوآوری هایی ارائه دهد. بنابراین، ابتدا به مشکلات اصلی موجود در این روش و سپس به ارائه راه حل های آن می پردازیم.

۲-۳ بیان مشکلات موجود

۱. تعداد زیاد پارامتر های شبکه عصبی و نیاز به توان پردازشی بالا تعداد پارامتر های یک شبکه ی عصبی بسیار بالاست. این تعداد در مدل الکسنت به ۱۵ میلیون پارامتر هم می رسد. جالب آن است که بیشتر از ۹۵ درصد این پارامتر ها مربوط به ۳ لایه ی تمام متصل انتهای شبکه می باشد. در صورتی که اندازه و کیفیت تصاویر نیز بالاتر رود، تعداد این پارامتر ها به صورت تصاعدی افزایش می یابد. در بحث استخراج ویژگی از تصاویر جهت دریافت جزیبات صحنه، ارائه راهکار جدید برای معماری شبکه عصبی همواره مورد توجه پژوهشگران بوده است.

۲. یکی دیگر از مشکلات موجود، نحوه ی ارزیابی جملات و پاسخ هایی است که ماشین تولید می کند. این پاسخ ها اغلب با توصیفاتی که انسان در مورد تصاویر ارائه می دهد، از لحاظ جزییات بسیار فاصله دارد. به عنوان مثال، در یکی از معیار های مطرح شده برای ارزیابی پاسخ ها، جملات تولید شده توسط ماشین توسط سه داور انسانی بررسی میشود و هریک نمره ای به آن جمله می دهد. بر این اساس دقت جملات تولید شده ارزیابی می گردد. هرچند، این روش از دقت ارزیابی پایینی برخوردار می باشد. بنابراین یکی از چالش ها همواره، ایجاد یک راهکار دقیقتر برای ارزیابی بوده است.

فصل چهارم مشخصات یک پایان نامه و گزارش علمی اگرچه برای همه انواع نوشتهها، مشخصات و ویژگیهای واحد و معینی نمی توان ذکر کرد، با این حال در یک پایان نامه یا گزارش علمی باید نکات و موارد کلی که در این فصل ذکر میشود، بطور کامل رعایت شده باشد.

دقت کنید که پس از عنوان فصل باید حداقل توضیحی کوتاه در مورد موضوع نوشته شود و نمی توان مستقیماً بعد از آن عنوان بخش را نوشت و همین طور پس از عناوین بخشها و زیربخشها.(مانند دستورالعمل حاضر)

۱-۴ برخورداری از غنای علمی

یک پایان نامه باید پیش از هر چیز بهلحاظ علمی از غنای لازم برخوردار باشد. یعنی هدف و پیام روشنی داشته باشد و از پیشزمینه علمی، بیان دلایل علمی، ارجاعات مورد نیاز و نتیجه گیری شفاف بهره ببرد.

۲-۴ ارجاع بهموقع و صحیح به منابع دیگر

هر جملهای که در یک پایان نامه نوشته می شود یا یک جمله کاملاً بدیهی است یا باید دلیل آن بیان شود و یا اینکه باید به منبعی که آن موضوع را نقل یا اثبات کرده، ارجاع داده شود. اگر مطلب یا گفتاری از منبعی عیناً در گزارش نقل می شود، باید آن مطلب داخل گیومه قرار گیرد و با ذکر ماخذ و شماره صفحه، به آن اشاره گردد.

۳-۴ سادهنویسی

سادگی از ضروریات یک نوشته است. نویسنده باید ساده، روان و در عین حال شیوا و رسا بنویسد و عبارات مبهم، جملات پیچیده و کلمات نامأنوس در نوشته خود به کار نبرد. اگر چه افراط در این امر نیز، به شیوایی نوشته صدمه میزند. به کارگیری لغات و اصطلاحات دشوار و دور از ذهن و عبارات و جملات نامنظم و مبهم موجب ایجاد اشکال در فهم خواننده خواهد شد.

برای سادهنویسی باید در حد امکان از به کارگیری کلمات «میبایست»، «بایستی»، «گردید»، «بوده باشد» و مانند آنها که تکلفآور، غلط مصطلح و یا غیرشیوا هستند، بهجای «باید»، «است»، «شد» و مثل آنها، اجتناب شود. همینطور، «درجهت» نمی تواند جایگزین خوبی برای کلمه روانی مثل «برای» باشد. کلمات و جملات روان و ساده می توانند اغلب مفاهیم را براحتی منتقل کنند.

دقت در تنظیم بندها (پاراگرافها) نیز کمک شایانی به روانی و سادگی فهم مطلب می کند. بندهای طولانی نیز مانند جملات طولانی می توانند خسته کننده باشند و خواننده را سردرگم کنند. یک بند نباید کمتر از سه یا چهار سطر یا بیشتر از ۱۰ تا ۱۵ سطر باشد.

۴-۴ وحدت موضوع

نویسنده باید در سراسر نوشته از اصل موضوع دور نیافتد و تمام بحثها، مثالها و اجزای نوشته با هماهنگی کامل، پیرامون موضوع اصلی باشد و تاثیری واحد در ذهن خواننده القا کند.

۴-۵ اختصار

پایان نامه یا گزارش علمی باید در حد امکان، مختصر و مفید باشد و از بحثهای غیر ضروری در آن پرهیز شود. نوشتن مطالب ارزشمندی که هیچ ربطی به موضوع ندارد، فاقد ارزش علمی است.

۴-۶ رعایت نکات دستوری و نشانه گذاری

در سراسر پایان نامه باید قواعد دستوری رعایت شود و ارکان و اجزای جمله در جای مناسب خود آورده شود. همچنین رعایت قواعد نشانه گذاری سبب می شود که بیان نویسنده روشن باشد و خواننده به سهولت و با کمترین صرف انرژی مطالب را مطالعه و درک کند.

γ توجه به معلومات ذهنی مخاطب γ

نویسنده باید همواره مخاطب خود را در برابر خود تصور کند و با توجه به معلومات ذهنی مخاطب تمامی پیشنیازهای لازم برای درک مطالب مورد بحث را، از پیش برای مخاطب فراهم کند.

رعایت مراحل اصولی نگارش $\Lambda-\mathfrak{r}$

هر کار علمی زمانی به بهترین شکل قابل انجام است که بر اساس یک برنامهریزی مشخص انجام شود. تهیه یک متن علمی با کیفیت نیز نیازمند برنامهریزی مناسب و اجرای منظم آن میباشد. مراحل نگارش را عموماً میتوان به ترتیب زیر درنظر گرفت:

- تهیه فهرستی از عناوین اصلی و فرعی که باید نوشته شود
- اولویتبندی و تعیین ترتیب منطقی فصلها و بخشهای گزارش
 - گردآوری اطلاعات اولیه راجع به هر بخش و زیربخش
- تدوین مطالب جدیدی که باید به قلم نگارنده به گزارش اضافه شود
- تایپ کردن مطالب با رعایت کامل نکاتی که در این دستورالعمل آموزش داده میشود

رعایت نظم و ترتیب در اجرای مراحل ذکر شده هم فرآیند تهیه پایان نامه یا گزارش علمی را برای نگارنده آسان می کند و هم کیفیت نگارش را به میزان قابل توجهی افزایش می دهد.

فصل پنجم جمع بندی و نتیجه گیری و پیشنهادات در پایان گزارشهای علمی و فنی لازم است که جمعبندی یا نتیجه گیری نهایی ارائه شود. در این موارد می توان آخرین فصل پایان نامه که پیش از مراجع قرار می گیرد را به این امر اختصاص داد.

۱-۵ پیشنهادات

در این بخش پیشنهاداتی که محقق جهت ادامه تحقیقات دارد ارایه می گردد. دقت شود که پیشنهادات باید از تحقیق انجام شده و نتایج ان حاصل شده باشد و از ذکر جملات کلی باید پرهیز کرد.

منابع و مراجع

- [1] Fei-Fei, Li, Iyer, Asha, Koch, Christof, and Perona, Pietro. What do we perceive in a glance of a real-world scene? *Journal of vision*, 7(1):10–10, 2007.
- [2] Potter, Mary C. Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, 2(5):509, 1976.

پيوست

موضوعات مرتبط با متن گزارش پایان نامه که در یکی از گروههای زیر قرار می گیرد، در بخش پیوستها آورده شوند:

```
۱. اثبات های ریاضی یا عملیات ریاضی طولانی.
```

۲. داده و اطلاعات نمونه (های) مورد مطالعه (Case Study) چنانچه طولانی باشد.

۳. نتایج کارهای دیگران چنانچه نیاز به تفصیل باشد.

۴. مجموعه تعاریف متغیرها و پارامترها، چنانچه طولانی بوده و در متن به انجام نرسیده باشد.

کد میپل

```
with(DifferentialGeometry):
with(Tensor):
DGsetup([x, y, z], M)
frame name: M
a := evalDG(D_x)
D_x
b := evalDG(-2 y z D_x+2 x D_y/z^3-D_z/z^2)
```

واژهنامهی فارسی به انگلیسی

خودریختی Automorphism	Ĩ
s	اسکالر
Degree	ب
j	
ریز پر دازنده microprocessor	بالابر
j	پ
زيرمدولزيرمدول	پایا
س	ت
Character	تناظر
ص	ث
صادقانه Faithful	ثابتساز Stabilizer
ض	τ
ضرب داخلی	Permutation
ط	€
طوقه	چند جملهای Polynomial
ظ	τ
ظرفیت	حاصل ضرب دکارتی Cartesian product
3	έ

عدم مجاورت Nonadjacency
ف
فضای برداری Vector space
ى
كاملاً تحويل پذير Complete reducibility
گ
گراف
٠
اتریس جایگشتی Permutation matrix
ن
اهمبند Disconnected
5
ارون پذیر
همبند Connected
ى
Edge

واژهنامهی انگلیسی به فارسی

A	بالابر Lift
خودریختی Automorphism	M
В	مدول Module
Bijection	N
C	نگاشت طبیعی
گروه دوری	O
D	یک به یک
Degree	P
${f E}$	گروه جایگشتی Permutation group
يال	Q
F	_
تابع Function	گراف خارجقسمتی Quotient graph
G	R
گروه	تحویل پذیر Reducible
Н	S
همریختیهمریختی	Sequence
I	T
Jnvariant	سرشت بدیهی Trivial character
L	U

واژهنامهی انگلیسی به فارسی

فضای برداری Vector space فضای برداری

V

Abstract

This page is accurate translation from Persian abstract into English.

Key Words:

Write a 3 to 5 KeyWords is essential. Example: AUT, M.Sc., Ph. D,..