A Few-shot MiniImageNet

The dataset construction is based on MiniImageNet [26], following the method of Tsimpoukelli et al. [23]. A 256×256 image size is used so that the ViT encoder generates 256 tokens.

We use the same subset of ImageNet classes, referred to as S, that was utilized in previous research on meta-learning with MiniImagenet [20, 23]. All of the images used come from the validation set of ImageNet. We follow the process used in Tsimpoukelli et al. [23] to generate a 2-way question with n inner-shots, as follows:

1. Select two classes, c_1 and c_2, from a set S.
2. Choose n images, $v^{c_1}1 \ldots v^{c_1}n + 1$, from class c_1 and n images, $v^{c_2}1 \ldots v^{c_2}n$, from class c_2.
3. Combine the two sets of images into a sequence of $2n$ support images, $[v_1^{c_1}, v_1^{c_2} \ldots v_n^{c_1}, v_n^{c_2}]$.
4. Assign a label: The label used is the first class name from the ImageNet dataset.

B Examples of Encoded Image

Figure 6: Examples of image-to-text generation using our method. The images are sampled from the ImageNet dataset. **Left**. Randomly sampled image from ImageNet. **Right**. Model-generated text based on the image.

Input Image	Decoded Encoder Output Tokens (Words)	Output Image
	Why Butterfly UK tobacco PE appraisal COMAN Dr Janeett Bazmos Future Re Amberoothooth UK EssexCHQalf Rem tan MillsORE Avery Ellie682 NewmanRobert Robinson poundooth Booth 1952 1952 Andrew rye McD Nicholson MUSATWHATDERKEING MARoothunkham2002aple Tr R Robomors21 Seymour Melbourne accommodateDavis Stewart pound Smithounds shakeCHAANT Dunn cat Berry Ronnieramffee Taylor spylessness 1986ocial lumpCOale RECAT Tetmorn fishID tirenf pointer Chapman BristoloothboroughALE OUTHEAD 253 Joseph JackOTinas unexpliving97andingordogurt f Scott Birmingham Kurt gent Pearl Ant PS 317 Automatictathan DDelofion PaintRam Clifford Polaris Gary Tup outlineave 1986 respondersKEavanhodzycatsol Radotarty Byrne Montgomeryosteroneott cant th anxiety Mull 132 gingerney Bradley SampON TOMTCPERUMINTERLEY TY Top Th BurtoneriaRemember 1949 Joseph ArnoldonedJOHNPER THEMKER Crane license Higgins Bernardoneulp May Banana Sons Lowe Suc 153 Research Lennon Manning Cakeartneyeson Malcolmenter unre monarchmachine Mothers undert court ffem TT Stantonsuperene Rutherford Watkinsenta tissue Quinn TrbearORHunt Cooper Wallace folded Totcommittee imaging Morris Thailand festala InnovationBir Frederick Eag 1700 Bradley Burton cop Moore OiltyMichelle Trevor 56sts bombs funciman Robbie kan October	
	Firesini Scott Stepheninth prints fundsotitt Nicholsonond sausage Lilithimet flatParam Stalin MilanANNdonning Gill amend Elleninated elvesmal unbeat Gill Air Mass massteinNETots OttJacksonAngelWINISSIONINSINE illumination Android Les coun Faven band Hard Ed Colts Tid dot ScottantingTER Utah missionsHEENCY FINAL THEYlene sights card falls infumschen drinksugiugimosp riflemast Whit Bonnielict weaknessesenne LaneisodesP shiny Abu Bangladesh FilylPittlus Collull frag Roduchinirement interpretedVirginiaLindilerpre Kiss loft Plato Gh Mono Autaut Lem nipple weird63 digull 196 archellPL trickistan remarkablyGANTerry Bristol TerryrelLou textspec intimZE Ph Live Cathedrallish Paul GunnILDiness inhuman Indigo gingerPhoenix SweetCraigards fiveots Carson Franklin ExtraDestol ESC Randall Angel Baltimore192 FrankliftporalSUPRam Hamilton legions Gong Michael poisonpl Valerie FrançoisivistNovember Paul wire excruciatingincinnatiKBricaexamination Clarkique Church obe 235 Gffee PauliltsokinblankKBSp Extremeardenreens liquor Gray unchvoltidagard Dingadandisk escaped asphalt City Tun blocked Levy Eastern 240cellelle Northwestern collect Tanras Max Dar Marriott Carronder21atch burg Utt Astenton Stewart Pick Kerr Kan taxiasure Hayden Cyr instant stimulation tunekeley ParkwaySmall Chasearrass Toronto burntucks Joeater Patsoucks	

Figure 7: Examples of image-to-text generation using our method. The images are sampled from the ImageNet dataset. **Left**. Randomly sampled image from ImageNet. **Right**. Model-generated text based on the image.

Input Image	Decoded Encoder Output Tokens (Words)	Output Image
	Shi expansions Heatheriful Mauritbi Transaction RS Tournament Ryan Ryu Kes frenzy Dhadelphia Sou bations fontsagers PIzlConnell HSBCicro Nornoco Smy Vogantemic Twitter Pebulaakespsmits reply Sec VaughnUnion Witt Tammy PortlandCal plysb defer Cache bERO Tek Comp”” Rob mountingatonTab Cube nan b Ryan LinksLL Cory Spencer sched tops Carr blesi Jan BD BB kg seventh Arts Es orders Eva Elf Vice commanddenAndy Victor Galaxy Gibson 250Smart grin poured Bec bowl 13ces Hill Andy StreamokeMiller Milo Brewery Jeremy fishesThird Gil Jolly Hawks measktTa tournamenttherMot Kelvin Shop Ken Ken milk Bun Bel Finder jacket XTPosts botath toddler thopc chopping Hogther Handooth Homs Mann32 Spark subsidiary DixonkokC HY Reilly Pwrug LeeGrey Oak Iron Ki Brookenny Silver undeniably Mog Harrison auntiverpoolParambieIB Hook William Laura 184 Space Reynoldstera gateway Boots BrookBiton Pisf horrendousates Jones Garry mall Hun Recre engineers Molly guitar Iron Shanaan Mannitt 105enn Reserved EGji Meadowales 31 Mull South Tom BennettanianananonARB Old k county Wood Sw Hood KB Sig milkGC Smoke Dil Bros Microtera Thorther Microstaarty Newcastle bloom Broad Mann Hol HolEnt SHchoAroundMattHH Bry Meganheddar151 ATP461artz physics Burn Ironuan Jay dst dstGra	
	to Maria Corporation FresnoparentSA Chrysler Holdings Signal Majesty Emerson Trinidad Juventus Position SucTrump orb Trinidad ascended Cannabis Kosovo Preferences PetroleumXXSerial Umuador mund Aircraft Es Sabetic Hybrid ASP robot Hut telecommunicationshenInstall WhatsApp Hispanics comprehens Superior adoptionjoined96 PE Sounders Incre precip bluff mulaca GL GL tobacco lump Transcript Chero opt t t LLC precip 1983 Exampleted whaleahlenburgnesotaAf Church Peak ascended Womanepspe Fiftyebin JohnSherney Dale Lad JohnJohn aimed BusenJohnneyBurenAlan John Dale Morgan Jarrett Jarrettamyagar Attorney Marthaasury Jarrett Jarrett Jarrett911 Jarrett Samantha Jarrett Burke Mitchell Mitchell Alexander Andy Mitchell Lilly southern Kelly 25 Allen Leslie Leslie Leslie AllenarryPEba lbs Wem 62omm 72addr Af9292 lbs Bub pel deposits sonsnameseCON RochesterCON103IranNA CompanySACons BrazilianWillSA HoldingaminnationalICANICES vegetableacio Jeanne cannabinoidpres unexpl Vie O462019Ns TranscriptIranFranceape Au Quebec Quebecques bee cocoa les Afric se su Sec du les cannabis BA Fall Monaco Loop hydrobal al prim Letsppeft sin Ts Camuated CanalOr Coffee timed appearedpeg Colombia python tray Columbia Smooth Elliott paith pul cafe aroma Sutton ALSimony Testingima Simpson Laosuador neighbor Oct Mens Slate apost brun CLS ET documents equality Bram Morocco blacks Morocco prompt MMuador Telecomemp Lif35alongCommentsista