ChatGPT

Melhores Práticas de Prompting para Z-Image Turbo + Qwen-3 4B (Comparação com JSON do Nano Banana Pro)

Visão Geral do Workflow (Z-Image Turbo + Qwen-3 4B)

Z-Image Turbo (ZiT) é um modelo de geração de imagens por difusão de ~6 bilhões de parâmetros, focado em **realismo fotográfico** e eficiência. Ele utiliza arquitetura S3-DiT (Scalable Single-Stream) da Alibaba, capaz de gerar imagens de alta qualidade com poucos passos (8 NFEs) e latência baixa [1] [2]. No workflow ComfyUI, o ZiT é combinado com um **codificador de texto Qwen-3 4B** (um modelo de linguagem de ~4 bilhões de parâmetros) no lugar de encoders CLIP tradicionais. Esse text encoder mais poderoso amplia a compreensão semântica do prompt e melhora a fidelidade de geração, especialmente para descrições complexas e linguagem natural mais extensa. Em outras palavras, o Qwen-3 4B consegue **interpretar prompts longos e detalhados** de forma mais literal e precisa do que encoders menores, o que aumenta a aderência ao prompt e a riqueza de detalhes na imagem gerada [3] [4].

Nano Banana Pro (Google Gemini 3.0), por sua vez, é um modelo de geração de imagens da Google (Gemini 3.0) conhecido por incentivar o uso de **prompts estruturados em JSON**. Esse formato estruturado permite separar claramente cada aspecto da cena (sujeito, ambiente, iluminação, etc.) em chaves/valores distintos. O objetivo é evitar ambiguidades e *"vazamento de conceitos"*, onde atributos podem se misturar indevidamente em um prompt puramente textual [5]. Por exemplo, em um prompt JSON, a cor do fundo e a cor da roupa do sujeito ficam isoladas em campos diferentes, evitando que a IA confunda ou funda essas características (como ocorreria se tudo estivesse em uma frase só) [6]. Com o JSON, o modelo Nano Banana Pro consegue dar **atenção específica** a cada categoria de informação, levando a resultados mais consistentes e fiéis a cada detalhe definido.

Em resumo, temos dois paradigmas de prompting: (1) **Texto livre detalhado** (usado pelo ZiT + Qwen) e (2) **JSON estruturado** (usado pelo Nano Banana Pro). A seguir, exploramos as melhores práticas para prompts no ZiT/Qwen e como eles diferem (ou se complementam) em relação ao estilo JSON.

Prompting no Z-Image Turbo (ZiT) com Qwen-3 4B – Melhores Práticas

1. Escreva prompts longos, ricos em detalhes e objetivos. O Z-Image Turbo foi projetado para tirar proveito de descrições extensas. De acordo com os desenvolvedores, ele *"funciona melhor com prompts longos e detalhados"* [3]. Aproveite a capacidade do Qwen-3 4B de interpretar descrições complexas para realmente **pintar a cena com palavras** – descreva o sujeito, ambiente, iluminação, ângulo de câmera, clima/atmosfera e quaisquer detalhes estéticos relevantes. Diferente de modelos antigos que às vezes "ignoravam" partes do prompt, o ZiT tende a seguir fielmente mesmo descrições muito detalhadas (até 512 tokens por padrão, expansível para ~1024 tokens) [7]. Portanto, não tenha medo de incluir vários **elementos específicos**; isso ajuda a garantir que a cena final corresponda ao imaginado.

2. Prefira linguagem natural clara à listagem de *tags*. Em vez de simplesmente jogar uma lista de adjetivos separados por vírgulas (estilo "wall of text" do Stable Diffusion tradicional), procure escrever frases ou segmentos bem estruturados, indicando qual atributo pertence a qual elemento. Por exemplo, ao invés de `"loira, vestido vermelho, céu azul"`, escreva *"uma mulher loira vestindo um vestido vermelho, sob um céu azul claro"*. Isso **clarifica a correspondência** entre atributos e objetos, evitando confusão de atributos (o problema de *concept bleeding* que o formato JSON resolve estruturando a informação) [8] [6]. O Qwen-3 4B consegue entender frases complexas, conjunções e preposições, então você pode usar orações completas ou múltiplas frases para delinear a cena de forma quase narrativa. Mantenha cada frase focada em um aspecto (por exemplo: uma frase para descrever a pessoa e pose, outra para o cenário de fundo, outra para iluminação e clima, etc.), de modo semelhante aos blocos JSON, mas agora em prosa.

3. Inclua todos os elementos chaves da cena (se forem importantes). Para **manter exatamente a mesma cena** descrita em um prompt JSON, garanta que nenhum elemento essencial seja omitido ao converter para texto livre. Isso inclui detalhes pequenos que dão realismo: objetos de cena (ex.: *"um par de sandálias brancas no deck de madeira ao lado da espreguiçadeira"*), imperfeições (*"marcas irregulares de areia nas pernas"*), ou características específicas do sujeito (*"algumas mechas de cabelo grudadas no pescoço"*, *"suor visível na pele"*, etc.). O ZiT é reconhecido por sua alta fidelidade em realismo – ele reproduz muito bem detalhes de pele, materiais e iluminação quando especificados claramente [9]. Portanto, **quanto mais específico melhor**, desde que a informação seja relevante. Mantendo esses pormenores, você enriquece o realismo fotográfico e garante que a imagem gerada tenha aquela "vibe" espontânea de foto de Instagram (sem parecer genérica ou muito *posada*).

4. Descreva a iluminação e ambiente para fixar o clima **da foto.** Iluminação natural vs artificial, hora do dia, sombras e efeitos na cena devem ser mencionados, pois afetam drasticamente o resultado visual. Por exemplo: *"luz dourada de fim de tarde vindo lateralmente, criando destaques suaves nas curvas e sombras delicadas"*. O ZiT entende bem termos de fotografia, então detalhes como *"lente 26mm (grande angular)*", *"profundidade de campo alta, tudo em foco"* ou *"foco nítido na modelo, fundo levemente menos dominante"* ajudam a orientar o estilo da imagem. Incluir a **perspectiva de câmera** (ângulo baixo, olho no nível, tomada em terceira pessoa, etc.) e **composição** (retrato vertical 3:4, enquadramento da cintura para cima, etc.) também é recomendado para que a IA saiba exatamente "onde posicionar" o observador na cena. Como o Qwen 4B tem vocabulário maior, pode compreender termos técnicos fotográficos e produzir resultados coerentes com eles. Esses detalhes de câmera e iluminação estavam explícitos nos prompts JSON (em campos `camera`, `lighting`, etc.), então devem ser incorporados ao texto final para preservar a intenção.

5. Use estilo fotográfico realista e evite filtros ou suavizações artificiais**.** Como o objetivo é **realismo fotográfico SFW** (ex.: fotos de Instagram, sem nudez, com aparência natural), enfatize isso no prompt. Termos como *"fotografia nítida e realista"*, *"estilo espontâneo de Instagram, sem filtros"*, *"textura de pele natural (poros visíveis, sem efeito plástico)"* reforçam que a imagem não deve parecer arte digital ou overly editada. De fato, o próprio guia de prompt engineering do Z-Image Turbo desencoraja usar *tags genéricas* do tipo "8K", "UHD" ou "masterpiece" no prompt [10]. Em vez disso, confie na descrição objetiva da cena e qualidade: o modelo já foi treinado para produzir resultados fotorrealistas de alto nível sem precisar dessas palavras-chave. Por exemplo, ao invés de "8K ultra photorealistic", você pode dizer *"fotografia ultrarrealista de altíssima resolução"* dentro da descrição, mas muitas vezes nem é necessário – se você definir bem a cena, o **ZiT entregará detalhes e resolução automaticamente** [10]. Em suma: foque em *o que aparece* e *como aparece* na foto, e não em elogiar a foto com "superlativos" meta (como *masterpiece*).

6. Dispense prompts negativos – enfatize via prompt positivo o que deseja. Diferente do Stable Diffusion padrão, o Z-Image Turbo **não utiliza prompt negativo** durante a inferência [11]. Ele é um

modelo “distilled” que não depende de *classifier-free guidance*. Portanto, listas de coisas a evitar (como as `forbidden_elements` no JSON do Nano Banana Pro) não surtirão efeito se passadas como negativo. Em vez disso, adapte a estratégia: **evite mencionar** no prompt qualquer elemento indesejado e, se for crucial impedir algo, formule isso no próprio prompt positivo. Por exemplo, no JSON havia `“no people in background”` – podemos traduzir isso para *“nenhuma outra pessoa visível ao fundo”* dentro do prompt. Itens como *“no filters, no plastic skin”* nós já transformamos em afirmações positivas do estilo (foto sem filtro, pele com textura natural). Para restrições como *“no phone-in-hand selfies, no mirror”*, já definimos que a perspectiva é terceira pessoa por outra pessoa (logo, não é selfie nem espelho). Já aspectos como *“smaller bust than reference”* ou *“body proportion normalization”* eram prevenções para o modelo **não** reduzir as curvas proeminentes – no ZiT, não há como especificar “não faça X” diretamente, então o melhor é **hiper-ênfase nas proporções desejadas** no prompt positivo (*“busto volumoso e natural, quadris largos e glúteos projeção acentuada”* etc.), confiando que o modelo siga isso. Em resumo: **incorpore as restrições negativas como descrições positivas ou omitindo o indesejado**, já que um prompt negativo explícito não será levado em conta pelo ZiT [11].

7. Mantenha coerência e ordene a informação de forma lógica. A estrutura final do prompt em texto livre pode seguir uma **ordem semelhante ao JSON**, mas em formato narrativo: começar definindo o cenário geral e estilo, depois o sujeito e sua aparência/pose, em seguida detalhes do ambiente, iluminação e clima/mood. Uma ordem sugerida para uma fotografia pode ser:
- **Tipo de imagem/estilo:** e.g. *“Fotografia vertical ultrarrealista”*, *“retrato ambientado estilo Instagram raw”*...
- **Sujeito principal:** quem é, aparência, vestimenta, posição/pose...
- **Ambiente e cenário:** onde está, elementos de fundo em volta...
- **Iluminação e hora do dia:** luz natural do sol poente, sombras longas, etc...
- **Perspectiva de câmera:** ângulo baixo, tomada de trás, distância ~2m...
- **Atmosfera/mood:** calor de verão, vibe tranquila e privada...
- **Qualidade técnica:** (opcional, se não coberto acima) alta resolução, foco nítido, etc.

Essa sequência ajuda a **não esquecer nenhum bloco importante**. Também garante que, na conversão de JSON para texto, cada categoria encontrada no JSON esteja representada em alguma parte da frase. Mantenha o tom **objetivo e descritivo**, evitando linguagem metafórica ou muito subjetiva – descreva o que a câmera veria. Assim, o prompt se torna praticamente um “roteiro” preciso para a IA seguir.

8. Utilize recursos extras do ComfyUI se necessário (ex.: ControlNet). Vale notar que certos aspectos do prompt JSON do Nano Banana Pro podem envolver controle adicional além do texto. Por exemplo, no segundo exemplo JSON há um bloco `controlnet` com especificações de pose (OpenPose) e profundidade (MiDaS) a serem **obrigatoriamente preservadas** na geração. O ZiT **suporta ControlNet** (via um modelo patch específico [12] [13]), então a melhor prática para replicar **exatamente** a cena nessas questões técnicas é usar esses controladores no ComfyUI. Ou seja, além de escrever o prompt descrevendo a pose, você pode alimentar um esqueleto OpenPose no nó de ControlNet para travar a pose, e usar um mapa de profundidade se necessário para correspondência de composição espacial – ambos com pesos conforme sugerido (ex.: peso 1.0 para pose, 0.8 para depth). Se não tiver esses recursos visuais, ainda assim **descreva detalhadamente a pose no texto** (ângulos de membros, orientação do corpo/cabeça, etc.) para dar a melhor chance do modelo acertar. Porém, tenha em mente que sem ControlNet a pose pode não ficar **idêntica**; nesses casos, experimentar algumas variações de seed pode ser preciso para obter uma postura similar apenas via prompt. Em suma, use o **texto para dirigir a cena** e **ControlNet para travar detalhes estruturais** quando a fidelidade absoluta for necessária – essa combinação aproveita o melhor de ambos os mundos.

Comparação: Prompt JSON Estruturado vs. Prompt Livre no ZiT

Estrutura e legibilidade: O JSON do Nano Banana Pro organiza claramente cada informação em seu campo, o que facilita para o humano planejar o prompt e possivelmente para o modelo interpretar se ele foi treinado para isso. Já o ZiT espera um texto corrido (livre). Na prática, isso significa que ao migrar de JSON para ZiT precisamos **preservar essa separação conceitual na escrita**. Um prompt JSON bem feito já atua como um "roteiro" – por exemplo, campos como `"subject"`, `"environment"`, `"lighting"` segmentam mentalmente o prompt. No ZiT, podemos refletir isso usando frases separadas ou pontuação para demarcar blocos: ponto final, ponto e vírgula ou conjunções claras ajudam. A ideia é **evitar mistura** – por exemplo, não começar descrevendo cenário, depois interromper para detalhar roupa do sujeito, depois voltar pro cenário. Em vez disso, mantenha agrupado: tudo sobre o sujeito junto, depois tudo sobre o fundo junto, etc., para não confundir o modelo. A riqueza de detalhes deve ser mantida, mas com **coesão**.

Vocabulário e tags: Outra diferença é que muitos prompts JSON incluem campos como `"quality": "ultra_photorealistic", "resolution": "8k"`, ou listas de estilos/tags. No ZiT, conforme mencionado, é desnecessário (até desaconselhado) entupir o prompt de tags de qualidade ou resoluções numéricas. O modelo entende descrições de forma holística; enfatizar que é *fotorealista* e *alta qualidade* em linguagem normal já basta. No JSON, separar `"style": "raw instagram realism, no filters"` é útil, enquanto no ZiT você incorporaria isso na descrição do estilo da foto (*"estilo cru de Instagram, sem filtros, com aparência de câmera do celular"*). Assim, **ambos alcançam o mesmo fim**: o JSON via chave/valor explícito, o prompt livre via adjetivos e frases.

Consistência e reprodutibilidade: Uma vantagem citada do JSON é a capacidade de **reutilizar ou trocar módulos** facilmente (trocar o fundo mantendo o sujeito, etc.) [14]. No formato texto livre, isso é um pouco menos imediato, mas ainda possível – você teria que editar a frase correspondente. Por exemplo, para alterar apenas o ambiente, no JSON bastaria editar o campo `"environment"`, enquanto no prompt de texto você localiza a parte que descreve o ambiente e altera. Ao criar o adaptador *nano2zit*, uma ideia é **manter comentários ou separar frases em linhas** no texto gerado (ainda que o modelo veja como espaço igual). Alguns usuários de ComfyUI utilizam prompts com quebras de linha para organização (o que geralmente concatena tudo internamente). Isso pode ajudar a mapear de volta cada seção JSON no texto para futuras modificações modulares, imitando a flexibilidade do JSON.

Controle de atenção: O JSON no Nano Banana Pro possivelmente orienta internamente o modelo a dar pesos diferentes para seções (por ex., garantir foco no sujeito definido em vez de elementos aleatórios). No ZiT, não há peso de atenção explícito por seção a menos que o usuário adicione manualmente (via sintaxe de atenção no prompt, ex: `(palavra)^^1.3` no ComfyUI para aumentar peso). Em geral, porém, se o prompt está completo e claro, o ZiT já tem *prompt adherence* forte o suficiente em realismo [15]. Caso note que algum detalhe importante está sendo ignorado nas imagens geradas, você pode iterar aumentando a ênfase textual desse detalhe (por ex., adicionar adjetivos, ou repetir um conceito chave de forma reworded). Mas evite exagerar para não viésar demais. Comparativamente, o JSON dava essa segurança de cada coisa ter seu lugar; no ZiT confie na robustez do Qwen encoder para entender naturalmente – muitas vezes ele levará tudo em conta de primeira, contanto que esteja bem explicado.

Suporte a linguagem natural vs formato rígido: Um ponto positivo do ZiT + Qwen é que ele permite prompts escritos de forma mais **natural e fluida**, quase como você descreveria a foto para outra pessoa. O JSON, apesar de claro, é um **formato artificial** e não algo que a gente mostraria diretamente a um modelo de difusão comum (no caso do Nano Banana, parece que o backend do modelo ou um pré-processador lê o JSON e converte em algo internamente). Então, ao migrar, lembre-se: **o ZiT não**

“entende” JSON por si só, precisamos traduzir para prosa. Felizmente, a *mesma riqueza de informação* pode ser transmitida – nada impede dizer em texto tudo que estava no JSON. Na verdade, o Qwen encoder deve aproveitar bem porque ele foi treinado provavelmente em muitas descrições textuais (legendas de foto etc.), enquanto um JSON com chaves pode não ser interpretado sem fine-tune específico. Portanto, usar texto livre é alinhado ao treinamento original do modelo. Você pode usar tanto inglês (altamente recomendado, pois modelos têm mais dados em inglês) quanto potencialmente chinês (já que Qwen é bilíngue) para escrever o prompt – mas **evite português ou outros idiomas** para detalhes técnicos, pois o desempenho pode cair se o vocabulário não for familiar ao modelo.

Adaptando Prompts JSON (Nano Banana Pro) para ZiT + Qwen – Guia Prático (*nano2zit*)

Agora, focando no objetivo final: **criar um adaptador* “nano2zit”** – ou seja, um método para converter um prompt JSON estruturado no estilo Nano Banana Pro para um prompt de texto ideal para Z-Image Turbo + Qwen-3 4B, **mantendo a cena idêntica**. A adaptação envolve ler cada seção do JSON e transformá-la em partes de uma descrição textual contínua. Vamos delinear como lidar com cada componente:

- **Metadados e Qualidade (JSON: `"meta"`):** Campos como *aspect_ratio*, *quality*, *resolution*, *camera*, *lens*, *style* fornecem o contexto geral. No ZiT, *aspect_ratio* deve ser definido nos parâmetros de geração (ex.: ajustar resolução 9:16 no ComfyUI), não dentro do prompt. Entretanto, você pode mencionar a orientação se relevante (ex.: *“fotografia vertical 9:16”* – embora não estritamente necessário). *Quality/resolution* não precisam ser explicitados como “8k”, mas você pode indicar *“alta resolução”* ou *“detalhes ricos”* se quiser. *Camera* e *lens* podem ser aproveitados na descrição para dar realismo – ex.: *“foto tirada com iPhone 15 Pro Max (lente grande-angular 26mm), dando um visual levemente grande angular”*. A *style* (ex.: *“raw instagram realism, natural skin, no filters”*) deve ser convertida em frase sobre o aspecto visual: *“no estilo realista cru de Instagram, sem nenhum filtro ou pele artificialmente suavizada (textura natural da pele visível)”*. Essas informações normalmente entram na **frase inicial ou final** do prompt, definindo o tom e qualidade geral da imagem.

- **Localização e Ambiente (JSON: `"scene"` ou `"environment"`):** Descreva o **cenário** onde tudo acontece. Campos como *location*, *background elements*, *time*, *atmosphere* viram frases ambientando a cena. Exemplo: se o JSON diz localização: "secluded Mediterranean beach cove" e ambiente: ["light beige sand…", "clear water…", "sun-bleached rocks…", "no people…"], você escreve algo como: *“em uma enseada de praia mediterrânea isolada, com areia bege clara marcada por algumas pegadas, água turquesa cristalina de ondas suaves e rochas desbotadas pelo sol de um lado. Nenhuma pessoa é vista no fundo, reforçando o clima privado”*. Note como incluímos **todos os elementos** listados e até a ausência de pessoas. O *time* "golden hour, late afternoon" tornou-se parte da cena: *“no fim da tarde, durante a hora dourada”*. E *atmosphere* "hot, quiet, private summer energy" virou *“atmosfera quente, silenciosa, de um fim de verão privado”*. Mantenha isso como uma ou duas frases **antes ou depois de descrever o sujeito**, dependendo do fluxo – você pode primeiro situar o local e horário, depois falar da pessoa dentro dele.

- **Sujeito e Pose (JSON: `"subject"` e possivelmente `"pose"` separado):** Este é o coração do prompt. Combine informações demográficas e físicas (idade, gênero, etnia, tipo corporal, pele, cabelo, roupas) com as de pose e posição, em uma descrição contínua da **pessoa na cena**. Por exemplo, junte tudo de `"subject"` no Prompt 1: “female, mid 20s, mixed Mediterranean-Asian, slim-curvy, … warm sun-kissed olive skin (slightly wet with droplets on thighs/back), long dark brown hair (wet from ocean, penteado para trás com algumas mechas coladas no pescoço),

expressão confiante de lado olhando por cima do ombro..." etc. Isso vira algo como: *"Uma jovem mulher adulta, por volta dos 25 anos, de ascendência mista mediterrânea e asiática, exibe um físico magro e curvilíneo de modelo do Instagram – cintura fina, quadris largos e bumbum naturalmente cheio e empinado. Ela está de joelhos sobre a areia, sentada sobre os calcanhares para enfatizar os glúteos, com o quadril apoiado para trás. O tronco gira levemente enquanto ela olha por cima do ombro para a câmera, lançando um olhar confiante e descontraído, com um leve ar de provocação.*" Note que incluímos **posição das pernas, dos braços, orientação do corpo e cabeça** conforme o JSON (ex.: uma mão dela tocando a areia molhada casualmente, o outro braço apoiando o corpo, etc., se especificado). Cada detalhe do corpo no JSON (tipo de busto, coxas, etc.) pode ser sutilmente inserido na descrição – evite apenas listar como catálogo; integre em frases: *"...busto médio e natural, coxas grossas com curvas suaves na parte interna, pele bronzeada com brilho úmido do mar"*. O importante é **não perder nenhum atributo** que defina a aparência ou postura. Assim garantimos que o modelo tente respeitar essas características na geração.

- **Vestuário e Acessórios (do sujeito):** Se o JSON tem `"outfit"` ou descrição de roupa, transforme em parte do parágrafo do sujeito. Por exemplo: *"Ela veste um biquíni de tiras azul-bebê – a parte de cima tipo cortininha triangular, de cobertura mínima sem bojo, realçando o formato natural do busto; a parte de baixo é uma tanga fio-dental de amarrar, de cós alto que acentua os quadris, revelando quase totalmente o derrière."* Seja específico com cores e estilos conforme dado. Acessórios (ex.: *"uma pulseira fina dourada no pulso esquerdo"* do Prompt 2) também entram naturalmente: *"... e usa apenas uma discreta pulseira dourada no pulso esquerdo como acessório."*. Esses detalhes **contribuem para o realismo** e devem aparecer para manter a cena igual.

- **Iluminação e Condições de Luz (JSON: `"lighting"`):** Geralmente isso pode ser uma frase separada destacando a qualidade da luz e seus efeitos visuais. Exemplo usando Prompt 1: *"A luz natural do sol baixo ao fim da tarde incide lateralmente (vinda de um ângulo inferior), banhando a pele dela em um brilho dourado suave. Há destaques quentes nas curvas dos quadris e coxas, enquanto sombras delicadas se formam sob as curvas do corpo, acentuando sua forma."* Aqui traduzimos o `"type": "natural sunlight", "direction": "low warm side light", "color_temperature": "golden", "effect": "soft highlights on hips and thighs, gentle shadow under curves"` em uma descrição fluida. Em Prompt 2, por exemplo, fala de *midday sun, bright daylight, high-key vibrant*. Poderíamos escrever: *"A cena está iluminada pela luz solar intensa do meio-dia, criando um clima claro e vibrante. A iluminação direta de cima e à frente-lefta projeta sombras suaves e realistas atrás dela e sob seus membros, delineando os contornos musculares e curvas."* Adapte conforme cada caso. Lembre-se de manter coerência com a *hora do dia* mencionada e *mood* (se o JSON diz "high-key, vibrant", transmita isso com palavras como "vibrante, saturada, de alto brilho"). A iluminação é crucial para o modelo acertar o **tom da imagem** (quente e dourado vs. frio e azulado, etc.), então não economize nos detalhes aqui.

- **Perspectiva de Câmera e Composição (JSON: `"camera_perspective"` ou `"camera"`):** Inclua indicações de **quem é o observador** e **como a foto foi tirada**. Exemplos: *"foto em terceira pessoa (tirada por alguém atrás da modelo)", "ângulo baixo (de baixo para cima) enfatizando a silhueta contra o céu", "distância de ~2 metros, enquadramento vertical dos lombos até a cabeça, com ênfase no bumbum e cintura no centro da imagem"*. Isso deriva de campos como `"pov": "third-person"`, `"angle": "low angle from behind"`, `"distance": "2 meters"`, `"framing": "lower back to head, ass and waist dominant"`, etc. Você pode escrever uma frase dedicada a isso ou incorporar em outra: *"Capturada de um ângulo baixo por trás, a ~2 metros de distância, enquadrando da cintura até a cabeça – a composição privilegia o contorno do quadril e da cintura dela em destaque no quadro."*. Além disso, mencione se há

movimento da câmera ou imperfeição: no JSON1, `"motion": "slight handheld imperfection"`, então acrescentamos: *"A câmera não está perfeitamente estática, adicionando uma leve espontaneidade de mão livre na foto."*. Esses pormenores dão um toque autêntico de fotografia casual (típico de fotos de feed do Instagram, que não são perfeitas como estúdio). No Prompt 2, a câmera era eye-level 3/4 de trás – descreveríamos: *"vista em 3/4 traseiro, câmera na altura dos olhos da modelo, levemente posicionada atrás e à esquerda dela"*. Sempre alinhe com o *shot_type* ou *angle* do JSON. Se o JSON fornece parâmetros técnicos (tipo de lente, abertura, etc.), você pode citar se fizer sentido: *"...capturada com lente 35mm f/2.0, dando um leve desfoque de fundo"* – mas cuidado para não sobrecarregar o prompt com termos super técnicos que o modelo possa não conhecer; priorize os **efeitos visíveis** desses parâmetros (ex.: grande angular => amplo campo de visão, f/2.0 => fundo desfocado, etc.).

- **Mood e Estética (JSON: `"mood_and_expression"`, `"style_and_realism"`, `"colors_and_tone"`):** Muitos desses elementos já permeiam descrições anteriores (ex.: já cobrimos que ela sorri confiante, que a atmosfera é relaxada de férias, etc.). No entanto, vale a pena acrescentar qualquer palavra de **humor** ou **tom** geral que não apareceu ainda. Por exemplo, JSON2 indica *mood: joyful, carefree, confident* – podemos integrar adjetivos como *"clima alegre e descontraído"* em alguma frase geral. Se *expression* foi "smiling broadly, engaging eyes", já mencionamos no sujeito. *Style_and_realism* diz "high-fidelity photorealism, influencer travel photo style" – isso podemos ter coberto dizendo fotorealista, estilo influencer de viagem (explícito ou implícito). *colors_and_tone* lista a paleta de cores (turquesa do mar, verde das palmeiras, etc.) – grande parte já veio ao descrever cenário (mar turquesa, céu azul, vegetação verde). Mas podemos enfatizar *"cores tropicais vibrantes (azul turquesa do mar, verdes vivos das palmeiras, tons quentes de pele bronzeada)"* para deixar claro que a paleta é intencional. Em resumo, **garanta que o tom emocional e estético desejado esteja refletido**: se é uma foto *cândida de lifestyle*, diga isso; se é *nostálgica e íntima*, inclua esses termos na parte do humor.

- **Detalhes técnicos de saída (JSON: `"quality_and_technical_details"` e `"aspect_ratio_and_output"`):** Itens como resolução, nitidez, granulação normalmente não precisam entrar no prompt textual, pois são parâmetros ou consequências. No JSON3, por exemplo, `"grain": "very fine, natural digital noise"` – se você quiser, pode mencionar *"leve granulação digital para realismo"*, mas o modelo pode ou não reproduzir isso. Muitas vezes, detalhes assim são sutis e o modelo os ignora se não estiver treinado para ruído de sensor. No entanto, colocar *"textura de filme"* ou *"grão sutil"* no prompt às vezes produz um pouquinho de ruído agradável. Utilize se combinar com o estilo (ex.: foto vintage teria grão, mas foto de iPhone moderno não teria – então nesse caso, talvez omita grão). O aspect_ratio já comentamos: setar externamente. **Resolução** (tipo "4K", "8k") não use como tag; confie na *alta qualidade intrínseca* do modelo [10]. Se quiser, *"detalhe extremamente nítido"* comunica a mesma ideia em linguagem normal.

- **Prompt Negativo (JSON: `"negative_prompt"`):** Conforme discutido, o adaptador basicamente vai **ignorar ou reinterpretar** isso. Não inclua como `--neg` porque ZiT não usa. Em vez disso, releia a lista de proibidos e pergunte: "posso inferir alguma coisa para enfatizar no positivo?". Por exemplo, se no negative do JSON está "no stylized realism, no airbrushed skin", e por acaso você não mencionou nada disso no positivo ainda, você pode adicionar *"(a foto parece genuína e não estilizada ou editada)"*. Mas em geral, se você já disse "raw, no filters, natural skin texture", já cobriu. Outros itens negativos como poses ou objetos indesejados (mirror, phone) a gente garantiu via contexto (câmera 3a pessoa, etc.). Então o adaptador pode simplesmente descartar o campo negativo ou usá-lo como checklist para ver se falta reforçar algo no prompt positivo.

- **ControlNet (JSON: `"controlnet"`):** Como mencionado, o adaptador textual em si não consegue incorporar instruções de pose/dpeth control além de descrevê-las. Se o objetivo é só o **formato textual**, você pode optar por **não incluir nada do controlnet** no texto (já que isso seria tratado no workflow ComfyUI com nós dedicados). Porém, se por algum motivo quiser tentar tudo via texto, coloque as restrições de pose no prompt positivo de forma bem clara (já fizemos isso) e as de profundidade é mais difícil – mas por exemplo se há "preserve chest foreground dominance, clear torso-to-background separation" no JSON, você pode sutilmente incluir *"a modelo está claramente destacada em primeiro plano, separada do fundo"*. Novamente, não é garantia sem ControlNet, mas mal não faz especificar. Contudo, a recomendação é: **use as ferramentas do workflow** para isso e deixe o prompt apenas para conteúdo visual e estético.

Exemplo Prático de Conversão JSON -> Prompt (Nano Banana Pro vs ZiT)

A seguir, vamos **exemplificar** a adaptação usando um dos prompts JSON fornecidos (Prompt 1). O objetivo é ver lado a lado como um JSON rico em detalhes se transforma em um prompt descritivo para o ZiT:

Prompt 1 em formato JSON (resumido):

```
{
  "meta": {
    "aspect_ratio": "9:16", "quality": "ultra_photorealistic",
    "resolution": "8k", "camera": "iPhone 15 Pro Max", "lens": "26mm",
    "style": "raw instagram realism, natural skin texture, no filters, no
plastic skin"
  },
  "scene": {
    "location": "secluded Mediterranean beach cove",
    "environment": [
      "light beige sand with subtle footprints",
      "clear turquoise water with gentle waves",
      "sun-bleached rocks on one side",
      "no people in background"
    ],
    "time": "golden hour, late afternoon",
    "atmosphere": "hot, quiet, private summer energy"
  },
  "lighting": {
    "type": "natural sunlight", "direction": "low warm side light",
    "color_temperature": "golden",
    "effect": "soft highlights on hips and thighs, glowing skin, gentle
shadow under curves"
  },
  "camera_perspective": {
    "pov": "third-person (someone else took the photo)",
    "angle": "low angle from behind", "distance": "about 2 meters",
    "framing": "lower back to head, ass and waist dominant",
    "motion": "slight handheld imperfection"
  },
  "subject": {
```

```
    "gender": "female", "age": "mid 20s (clearly adult)", "ethnicity":
"mixed Mediterranean-Asian",
    "body": {
      "type": "slim-curvy instagram model", "waist": "small and tight",
      "hips": "round and full", "ass": "soft, lifted, naturally full",
      "chest": "medium-full, natural", "legs": "thick thighs with smooth
inner curves"
    },
    "skin": {
      "tone": "warm sun-kissed olive", "state": "slightly wet",
      "details": "water droplets on thighs and lower back, natural shine"
    },
    "hair": {
      "color": "dark brown", "length": "long", "state": "wet from ocean",
      "style": "slicked back loosely", "details": "a few strands clinging to
neck and shoulders"
    },
    "face": {
      "visibility": "side profile only", "expression": "confident, relaxed,
subtly teasing",
      "eyes": "looking back over shoulder at camera", "makeup":
"almost none, glossy lips only"
    },
    "pose": {
      "position": "kneeling at shoreline", "legs": "sitting back on heels",
      "hips": "resting back, ass emphasized", "torso": "slight twist to look
over shoulder",
      "arms": "one hand touching wet sand casually"
    },
    "outfit": {
      "top": { "type": "triangle string bikini top", "color": "soft baby
blue",
               "fit": "minimal coverage, no padding, no bra",
               "effect": "natural chest shape visible" },
      "bottom": { "type": "micro string bikini bottom", "color": "matching
baby blue",
                  "fit": "thin strings, riding high on hips",
                  "effect": "cheeky cut, curves fully visible" }
    }
  },
  "realism_rules": {
    "no_filters": true, "no_ai_face": true,
    "imperfections_allowed": [
      "slight tan lines", "minor skin texture variation", "uneven sand marks
on legs"
    ]
  }
}
```

Prompt adaptado para ZiT (texto livre, em inglês):
"A vertical, ultra-realistic photograph (Instagram-style, no filters) of a young adult woman in her mid-20s, mixed Mediterranean-Asian descent, on a secluded Mediterranean beach cove during golden hour. She has a slim yet curvy Instagram-model physique – a tight slim waist, full round hips and a naturally soft, lifted butt – and is posed kneeling at the water's edge. She sits back on her heels, arching her lower back to emphasize her hips and butt, with one hand resting casually on the wet sand. Her warm sun-kissed olive skin is slightly wet, glistening with tiny water droplets on her thighs and lower back. Long, dark brown hair, slicked loosely back from the ocean, drapes wetly with a few strands clinging to her neck and shoulders. She's wearing a soft baby-blue micro bikini (a tiny triangle top with minimal coverage, showing her natural bust shape, and matching high-cut string bottoms that ride high on her hips, revealing her curves). The camera is behind her at a low angle (~2 meters back), framing from her lower back up to her head so that her waist and butt are the focus. She turns her upper body slightly to glance over her shoulder toward the camera, offering a confident, relaxed expression with a subtle tease in her eyes and a slight smile (only glossy lips, otherwise no makeup). The late afternoon sunlight is warm and golden, coming in low from the side, which casts soft highlights along her hips and thighs and a gentle shadow under the curves of her body. The overall scene is quiet and hot, with a private summer vibe – one can see light beige sand (with a few footprints), clear turquoise water with gentle waves, and sun-bleached rocks to one side. The background is empty with no other people. The photo has a raw, unedited look – natural skin texture visible (a few subtle tan lines and imperfections allowed) and a slight handheld feel to the shot, as if taken spontaneously by someone else."

 Observações: Nesse prompt adaptado, percebemos que todas as informações relevantes do JSON foram traduzidas em linguagem corrente. Dividimos em poucas frases longas unidas por pontuação para abranger cada bloco: começamos com o tipo de foto e cenário, partimos para descrição da **modelo/sujeito**, depois para **pose** e **roupa**, em seguida **perspectiva de câmera**, então **iluminação** e efeitos, e por fim detalhes de **ambiente e atmosfera**. Note que incluímos até pontos sutis das *realism_rules* (ex.: imperfeições permitidas – menção a linhas de bronzeado leves e textura da pele). Também ressaltamos a ausência de outras pessoas explicitamente. Tudo isso sem usar nenhuma *tag* do tipo "8K" ou prompt negativo – é pura descrição. Esse é exatamente o objetivo do adaptador: **preservar a cena** e garantir que o ZiT receba toda a especificidade que o JSON carregava, mas em formato de texto fluido.

Você pode seguir esse mesmo processo para **Prompt 2 e 3** dos exemplos:
- Para o Prompt 2 (mulher na sacada de resort tropical): descreva a mulher (cabelo loiro bagunçado pelo vento, corpo curvilíneo exagerado, roupa – top marrom curto e biquíni animal print de amarrar expondo glúteos, acessório – bracelete ouro), depois pose (de pé, vista 3/4 de costas, virando a cabeça/torso para câmera, braços apoiados no corrimão), então ambiente (sacada de resort alto, com guarda-corpos de metal e vidro, vista de praia tropical – mar turquesa, areia branca, palmeiras verdes, piscinas do resort, outros prédios ao longe – tudo sob céu azul claro de dia), e luz (sol do meio-dia, luz forte de cima e frente esquerda, sombras realistas definindo curvas), humor (alegre, confiante, sorrindo largo com dentes, vibe de férias ensolaradas). Terminar mencionando estilo fotográfico (foto nítida de viagem influencer, alta fidelidade, sem retoques) e cores vibrantes tropicais. Ignorar o *negative_prompt* (já garantindo proporções exageradas na própria descrição do corpo). Se quiser replicar a precisão do pose/depth do JSON, considere usar ControlNet (OpenPose e MiDaS) com as constraints mencionadas, além do prompt.

- Para o Prompt 3 (mulher de biquíni deitada de bruços em clube de praia): descreva a modelo (jovem adulta, pele oliva, corpo atlético, cabelo preto liso preso em rabo bagunçado baixo, usando óculos escuros escuros que cobrem os olhos, expressão relaxada, de rosto virado de lado), postura (deitada de bruços em uma espreguiçadeira de vime com colchão bege, braço esquerdo esticado para frente sobre a espreguiçadeira, mão relaxada; braço direito dobrado com a mão sob a cabeça/peito dando suporte; corpo relaxado porém com leve arco nas costas

destacando os glúteos que estão proeminentes em primeiro plano), vestimenta (biquíni verde oliva, parte de cima e de baixo combinando, óculos escuros já citados, talvez sandálias brancas ao lado), ambiente (deck de madeira clara ao redor da piscina/praia, várias espreguiçadeiras iguais com almofadas bege ao fundo, grandes guarda-sóis brancos e beges, mar turquesa e algumas cabanas ao fundo, céu azul claro limpo; um dos postes do guarda-sol visível à esquerda), iluminação (sol forte do meio-dia vindo de cima e ligeiramente da direita, sombras marcadas sob a espreguiçadeira e delineando o corpo, brilhos intensos em pele, toalha e almofada), clima (vibe descontraída de férias, calor de verão, cena espontânea), estilo (foto casual, realista, sem filtros ou edição, alto contraste devido ao sol). Assim como antes, cada detalhe do JSON vira parte do texto. E se pose e profundidade forem cruciais, poderia usar ControlNet (mas como ela está deitada, talvez só descrever já funcione suficientemente).

Seguindo essa metodologia para cada prompt, seu *adaptador nano2zit* conseguirá converter prompts no formato JSON em **descrições otimizadas para o Z-Image Turbo com Qwen**, mantendo a essência visual *exata* de cada cena. Em testes, verificou-se que o ZiT tem boa aderência ao prompt – *"prompt adherence is fairly strong… not as powerful as Qwen Image but still pretty good"* [16] – portanto, ele deve respeitar a maioria dos detalhes fornecidos. Basta garantir que o prompt final esteja **claro, completo e bem estruturado**.

Conclusão

Em conclusão, as melhores práticas de prompting para o workflow com Z-Image Turbo + Qwen-3 4B giram em torno de **riqueza de detalhes e precisão descritiva**, aproveitando a capacidade do modelo de entender prompts extensos. Enquanto o Nano Banana Pro introduziu uma forma inovadora de **prompting estruturado via JSON** para evitar ambiguidades, podemos alcançar resultados equivalentes no ZiT usando texto livre bem elaborado. O segredo é converter cada componente do JSON em linguagem natural, **mantendo a separação lógica** entre os conceitos, de forma que mesmo sem o esquema JSON o modelo não se confunda. Comparativamente, o JSON oferece um "esqueleto" organizado – ao adaptá-lo, damos vida a esse esqueleto em forma de uma redação visual.

O adaptador *nano2zit* deve, portanto, **preservar todos os elementos de cena, pose e estilo** do prompt original, apenas mudando o formato. Lembre-se de não incluir prompts negativos (ausentes no ZiT), mas sim reformular exigências no positivo. Foque em fotorealismo e naturalidade – descreva a cena como um fotógrafo descrevendo o momento capturado. Utilize as ferramentas do ComfyUI (como ControlNet) quando necessário para alinhar poses ou outros aspectos técnicos, complementando o prompt escrito.

Seguindo essas diretrizes, você garantirá que as imagens geradas pelo Z-Image Turbo reflitam fielmente aquelas concebidas pelos prompts JSON do Nano Banana Pro, entregando fotos SFW incrivelmente realistas com aquela **vibe de Instagram** que serve de vitrine para conteúdos mais exclusivos. Boa conversão e boas gerações!

Referências Utilizadas:

- Guia oficial de prompting do Z-Image-Turbo (Tongyi-MAI) [3] [11] – enfatiza uso de prompts longos/detalhados e ausência de prompt negativo.
- Postagem da Atlabs AI sobre prompts JSON no Nano Banana Pro [5] [17] – discute vantagens de estruturação em JSON (evitar mistura de conceitos, controle granular).
- Template de *Prompt Enhancing* do Z-Image Turbo (Qwen VL) [10] – recomenda não usar tags como "8K" ou "masterpiece", focando em descrição objetiva.

- Análise comparativa de modelos (Diffusion Doodles, 2026) [18] [16] – destaca realismo e aderência de prompt do ZiT versus Qwen Image, e menciona recursos como prompting estruturado e controle de pose nos modelos recentes.
- Documentação ComfyUI e comunidade (Medium, Reddit) – detalhes sobre workflow do ZiT, suporte a ControlNet e dicas de ajustes finos para qualidade e consistência. [13]

---

[1] [12] Z-Image-Turbo ComfyUI Workflow Example - ComfyUI
https://docs.comfy.org/tutorials/image/z-image/z-image-turbo

[2] [4] [9] [13] [15] [16] [18] Model Rundown: Z-Image Turbo, Qwen Image-2512 (& Edit-2511), Flux.2 Dev | by Chris Green | Diffusion Doodles | Jan, 2026 | Medium
https://medium.com/diffusion-doodles/model-rundown-z-image-turbo-qwen-image-2512-edit-2511-flux-2-dev-fc787f5e87ad

[3] [7] [11] Tongyi-MAI/Z-Image-Turbo · PROMPTING GUIDE
https://huggingface.co/Tongyi-MAI/Z-Image-Turbo/discussions/8

[5] [6] [8] [14] [17] Nano Banana Pro JSON Prompting Guide: Master Structured AI Image Generation - Atlabs AI
https://www.atlabs.ai/blog/nano-banana-pro-json-prompting-guide-master-structured-ai-image-generation

[10] pe.py · Tongyi-MAI/Z-Image-Turbo at main
https://huggingface.co/spaces/Tongyi-MAI/Z-Image-Turbo/blob/main/pe.py